

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2004, 02:22:49 ; Search time 2482 Seconds
(without alignments)
10076.120 Million cell updates/sec

Title: US-09-867-753-1

Perfect score: 577

Sequence: 1 tccacatcagcgcgtccag.....ctacatcgcgtgagctag 577

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ha: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_scs: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_scs: *
28: em_un: *
29: em_vl: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcg_mus: *
34: em_hcg_pln: *
35: em_hcg_rtd: *
36: em_hcg_mam: *
37: em_hcg_vrt: *
38: em_hcg_hum: *
39: em_hcg_mus: *
40: em_hcg_hum: *
41: em_hcg_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	577	100.0	577	6	AX322486	AX322486 Sequence
2	566	98.1	755	6	AY099086	AY099086 Homo sapi
3	566	98.1	979	6	AR339190	AR339190 Sequence
C 4	409	70.9	10968	6	AX322488	AX322488 Sequence
C 5	409	70.9	170967	9	AC005023	AC005023 Homo sapi
C 6	399.4	69.2	139483	2	AC145687	AC145687 Pan trogl
C 7	383.4	66.4	204520	2	AC146356	AC146356 Pongo pyg
8	348.2	60.3	157119	2	AC146353	AC146353 Macaca mu
9	106.8	18.5	1205	9	HS421131	AL590524 Novel hum
10	105.8	18.3	1194	9	AK058125	AK058125 Homo sapi
11	105.8	18.3	1219	9	BC021719	BC021719 Homo sapi
12	105.8	18.3	1274	9	AF317219	AF317219 Homo sapi
13	105.8	18.3	1297	9	HS525N141	AL590526 Novel hum
14	103	17.9	1451	9	AY114148	AY114148 Homo sapi
15	103	17.9	1518	6	AX834967	AX834967 Sequence
16	103	17.9	1518	9	AK097704	AK097704 Homo sapi
17	103	17.9	1580	9	BC053599	BC053599 Homo sapi
C 18	102	17.7	157119	2	AC146353	AC146353 Macaca mu
19	97.2	16.8	1325	10	AF017734	AF017734 Mus muscu
20	97.2	16.8	1408	10	AF017735	AF017735 Mus muscu
21	97.2	16.8	1616	10	AF085715	AF085715 Mus muscu
22	97.2	16.8	1720	10	AF004211	AF004211 Mus muscu
23	90.6	15.7	175366	10	AC101718	AC101718 Mus muscu
24	74.8	13.0	4329	5	AF001393	AF001393 Oryzias l
25	73.8	12.8	1764	5	X1AF001049	AF001049 Xenopus l
26	73	12.7	868	10	BC049711	BC049711 Mus muscu
27	72.4	12.5	1913	3	AF165886	AF165886 Branchio
28	71.8	12.4	882	10	AF017453	AF017453 Mus muscu
29	71.8	12.4	1326	9	BC010923	BC010923 Homo sapi
30	71.8	12.4	1414	5	AB006104	AB006104 Danio rer
31	71.8	12.4	1442	9	HSU31986	U131986 Human carti
32	71.6	12.4	1719	5	AY130460	AY130460 Xenopus l
33	71.6	12.4	220851	2	AC097956	AC097956 Rattus no
C 34	71.2	12.3	286	10	AY116506	AY116506 Mus muscu
35	70.2	12.2	2368	5	X1AF001048	AF001048 Xenopus l
36	70	12.1	740	10	AY147207	AY147207 Mus muscu
37	70	12.1	1639	5	BC056324	BC056324 Danio rer
38	69.6	12.1	624	10	S82627	S82627 Rattus sp.
39	69.6	12.1	1494	5	AF121771	AF121771 Danio rer
40	69.2	12.0	1164	5	ZERGHDA	L03394 Brachydanio
41	69.2	12.0	1252	5	ZERGHDA	L03395 Brachydanio
42	68.6	11.9	1485	5	AF017273	AF017273 Xenopus l
43	68	11.8	1164	5	AF092538	AF092538 Gallus ga
44	68	11.8	2399	10	RN029174	U29174 Rattus norv
45	67.6	11.7	2032	5	PMA458324	AJ458324 Petromyzo

ALIGNMENTS

RESULT 1	AX322486	577 bp	DNA	linear	PAT 07-JAN-2002
LOCUS	AX322486	Sequence 1 from Patent EP1162265.			
DEFINITION	AX322486				
ACCESSION	AX322486.1	GI:18093553			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
	Homo sapiens (human)				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE					
AUTHORS	Haendler, B., Weiss, B. and Geserick, C.				
TITLE	Human pem as a target for fertility control and alzheimer therapy				
JOURNAL	Patent: EP 1162265-A 1 12-DEC-2001;				

Pred. No. is the number of results predicted by chance to have a

Schering Aktiengesellschaft (DB)

FEATURES
Location/Qualifiers
source

1..577
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 577; DB 6; Length 577;
Best Local Similarity 100.0%; Pred. No. 2e-113;
Matches 577; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCACATCAGGCGCTCCAGCCATGCGCGCTGCTGCTCCAGACACCGTGTCTACTG 60
DB 1 TCCACATCAGGCGCTCCAGCCATGCGCGCTGCTGCTCCAGACACCGTGTCTACTG 60
QY 61 CCTGAGTATATACAGAGTAAAAATAGCCCCACCTCAGCTGGGGGCGCATCAAGCCG 120
DB 61 CCTGAGTATATACAGAGTAAAAATAGCCCCACCTCAGCTGGGGGCGCATCAAGCCG 120
QY 121 AGAAGGCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTATATGAACTCTGAGGCGG 180
DB 121 AGAAGGCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTATATGAACTCTGAGGCGG 180
QY 181 TGTGAACCAAGAAACGCGATGAACCGCATGCGGCGCATGATCCCGAGGCGCGTGG 240
DB 181 TGTGAACCAAGAAACGCGATGAACCGCATGCGGCGCATGATCCCGAGGCGCGTGG 240
QY 241 AAAACGAGAGCTCCGCGACGACCGGACCGCGCGGAGAGGCGCGCCAGGCGGCGCAT 300
DB 241 AAAACGAGAGCTCCGCGACGACCGGACCGCGCGGAGAGGCGCGCCAGGCGGCGCAT 300
QY 301 GGAGGCTCCGCGACCGCGAAACATGAGCAGCACTCGGCGCGACGAGTTCAAGCTGTT 360
DB 301 GGAGGCTCCGCGACCGCGAAACATGAGCAGCACTCGGCGCGACGAGTTCAAGCTGTT 360
QY 361 GCAAGTGAAGAGAGCTGAAAGTGTTCGACACACTCAATACCTGATGTGCCAAG 420
DB 361 GCAAGTGAAGAGAGCTGAAAGTGTTCGACACACTCAATACCTGATGTGCCAAG 420
QY 421 AAGGGAATTCGCGAAACTGATAGTGTGACTGAAGCAAAAGTGGGTTGTTAAGAA 480
DB 421 AAGGGAATTCGCGAAACTGATAGTGTGACTGAAGCAAAAGTGGGTTGTTAAGAA 480
QY 481 TAAAGGCGCAGATGAGCGACATCAGAGAGATTAATGCTGGCCATGAATCACTGTC 540
DB 481 TAAAGGCGCAGATGAGCGACATCAGAGAGATTAATGCTGGCCATGAATCACTGTC 540
QY 541 TGAACCAAGAGAGTGTCTTCACTGCTGTGACTAG 577
DB 541 TGAACCAAGAGAGTGTCTTCACTGCTGTGACTAG 577

RESULT 2

AY099086

LOCUS Homo sapiens paired-like homeobox protein OTEX mRNA, complete cds.
DEFINITION Homo sapiens paired-like homeobox protein OTEX mRNA, complete cds.
ACCESSION AY099086
VERSION AY099086.1 GI:20563000

KEYWORDS Homo sapiens (human)
SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Geserick, C., Weiss, B., Schleuning, W.D. and Haendler, B.
TITLE OTEX, an androgen-regulated human member of the paired-like class of homeobox genes

JOURNAL Biochem. J. 366 (Pt 1), 367-375 (2002)

MEDLINE

22152918

PUBMED

11980563

REFERENCE

2 (bases 1 to 755)

AUTHORS Geserick, C., Weiss, B. and Haendler, B.

TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) Experimental Oncology, Schering AG,
Muelletstr. 170-178, Berlin 13342, Germany

FEATURES
Location/Qualifiers
source

1..755
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xq24"

ORIGIN

CDS

Query Match 98.1%; Score 566; DB 9; Length 755;
Best Local Similarity 100.0%; Pred. No. 4.5e-111;
Matches 566; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

QY 12 GCGCTCCAGCCATGCGCGCTGCTGCTCCAGACACCGTGTCTACTGCTGAGTGTAT 71
DB 65 GCGCTCCAGCCATGCGCGCTGCTGCTCCAGACACCGTGTCTACTGCTGAGTGTAT 124
QY 72 ACCAGGTAAATAATAGCCCAACCTCAGCTGGGGGCGAGCATCAAGCGAAGAGCGATG 131
DB 125 ACCAGGTAAATAATAGCCCAACCTCAGCTGGGGGCGAGCATCAAGCGAAGAGCGATG 184
QY 132 TTGGCCAAAGAGCTCCAGGCTCATGGGTATATGAACTCTGAGGCGGTGTGAACCAAG 191
DB 185 TTGGCCAAAGAGCTCCAGGCTCATGGGTATATGAACTCTGAGGCGGTGTGAACCAAG 244
QY 192 AGAAGGCATGAACCGCATGAGCGGCGATATCCCGAGGCGCGGCTGGAACCAAGAGC 251
DB 245 AGAAGGCATGAACCGCATGAGCGGCGATATCCCGAGGCGCGGCTGGAACCAAGAGC 304
QY 252 CTCGCGACGACCGCGACGCGCGCGCGGAGAGACCGCGCGCGCATGAGAGGTCGCCG 311
DB 305 CTCGCGACGACCGCGACGCGCGCGCGGAGAGACCGCGCGCGCATGAGAGGTCGCCG 364
QY 312 AGCCGAGAAATGACGACGAGCACTCGGCGACGAACTGTCAGCTGTGCGAGGTGAAG 371
DB 365 AGCCGAGAAATGACGACGAGCACTCGGCGACGAACTGTCAGCTGTGCGAGGTGAAG 424
QY 372 AGCTGGAAGTGTTCGACACACTCAATACCTGATGTGCCACAGAGGAACTTG 431
DB 425 AGCTGGAAGTGTTCGACACACTCAATACCTGATGTGCCACAGAGGAACTTG 484
QY 432 CCGAAACTTAGTGTGATGAAGCAAAAGTGGGTTGTTAAGATTAAGAGGCCCA 491
DB 485 CCGAAACTTAGTGTGATGAAGCAAAAGTGGGTTGTTAAGATTAAGAGGCCCA 544
QY 492 GATGTAGCGGACATCAGAGAAATTAATGCTGGCCATGAATCACTGTCGAGCCAGAG 551
DB 545 GATGTAGCGGACATCAGAGAAATTAATGCTGGCCATGAATCACTGTCGAGCCAGAG 604
QY 552 ACTGTGTCTTCACTGCTGTGACTAG 577
DB 605 ACTGTGTCTTCACTGCTGTGACTAG 630

RESULT 3

AR339190

LOCUS AR339190 979 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 681 from patent US 6569662.
ACCESSION AR339190
VERSION AR339190.1 GI:33726047

KEYWORDS

Source	Organism	Reference	Authors	Title	Journal	Features	Source
Unknown.	Unknown.	Unclassified.	1 (bases 1 to 979)				
REFERENCE			Tang, Y.T., Zhou, P. and Drmanac, R.T.				
AUTHORS			Nucleic acids and polypeptides				
TITLE			Patent: US 6596662-A 681 27-MAY-2003;				
JOURNAL			Location/Qualifiers				
FEATURES			1..979				
SOURCE			/organism="unknown"				
ORIGIN			/mol_type="genomic DNA"				
Query Match		98.1%;	Score 566;	DB 6;	Length 979;		
Best Local Similarity		100.0%;	Pred. No. 4.5e-111;				
Matches	566;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;
QY		12	GCAGCTCCAGCCATGAGCGCGCTTCGCTCCGTCACAGACACCGTGTTCACCTGCTGAGTAT	71			
DB		412	GCAGCTCCAGCCATGAGCGCGCTTCGCTCCGTCACAGACACCGTGTTCACCTGCTGAGTAT	471			
QY		72	ACCAAGTAAATATTAAGCCCAACCTCAGCTGAGGAGCATCAAGCGCAGAGGCCATG	131			
DB		472	ACCAAGTAAATATTAAGCCCAACCTCAGCTGAGGAGCATCAAGCGCAGAGGCCATG	531			
QY		132	TTGGCCAAAGAGCTTCAGGCGCTCATGGGTAATTTGAACCTTGAGGGCGGTGTGAACCAAG	191			
DB		532	TTGGCCAAAGAGCTTCAGGCGCTCATGGGTAATTTGAACCTTGAGGGCGGTGTGAACCAAG	591			
QY		192	AGAACGGGATGAAACCGGATGAGCGGATGATCCCGAGAGGCGGCGGTGTGAACCAAGAGC	251			
DB		592	AGAACGGGATGAAACCGGATGAGCGGATGATCCCGAGAGGCGGCGGTGTGAACCAAGAGC	651			
QY		252	CTCGGACAGCAGCCGCGAGCCCGCGCGAGAGAGCCGCGCCAGCGGCGCATGAGAGGTCGCG	311			
DB		652	CTCGGACAGCAGCCGCGAGCCCGCGCGAGAGAGCCGCGCCAGCGGCGCATGAGAGGTCGCG	711			
QY		312	AGCCCGAGAACATGACGACCAAGAACTCGGCGGACGAAGTTCAACGCTTTGCAAGTGAAGG	371			
DB		712	AGCCCGAGAACATGACGACCAAGAACTCGGCGGACGAAGTTCAACGCTTTGCAAGTGAAGG	771			
QY		372	AGCTGGAAAGGTTTTCCGACACACTCAATTAACCCGATGTGCGCCACAAGAAAGGAACTTG	431			
DB		772	AGCTGGAAAGGTTTTCCGACACACTCAATTAACCCGATGTGCGCCACAAGAAAGGAACTTG	831			
QY		432	CCGAAACTTATAGTGTGACTGAAAGCAAGTGCGGGTTTGTTTAAGATTAAGAGGCCA	491			
DB		832	CCGAAACTTATAGTGTGACTGAAAGCAAGTGCGGGTTTGTTTAAGATTAAGAGGCCA	891			
QY		492	GATGTAGGCGACATCAAGAGAAATTAATGCTTCGCCAATGAATCAAGTCTGACCCAGAGC	551			
DB		892	GATGTAGGCGACATCAAGAGAAATTAATGCTTCGCCAATGAATCAAGTCTGACCCAGAGC	951			
QY		552	ACTGTGTCTACATGCTGCTGGAAGTCTG	577			
DB		952	ACTGTGTCTACATGCTGCTGGAAGTCTG	977			
RESULT 4							
LOCUS	AX322488/c						
DEFINITION	Sequence 3 from Patent EP1162265.						
ACCESSION	AX322488						
VERSION	AX322488.1						
KEYWORDS	GI:18093554						
SOURCE							
ORGANISM	Homo sapiens (human)						
REFERENCE							
AUTHORS	Haendler, B., Weiss, B. and Geserick, C.						
TITLE	Human pem as a target for fertility control and alzheimer therapy						
JOURNAL	Patent: EP 1162265-A 3 12-DEC-2001;						

FEATURES	Schering Aktiengesellschaft (DE)	Location/Qualifiers
source	1. 10968	/organism="Homo sapiens"
		/mol_type="unassigned DNA"
		/db_xref="taxon:9606"
ORIGIN		
Query Match	70.9%; Score 409; DB 6; Length 10968;	
Best Local Similarity	100.0%; Fred. No. 1.7e-77;	
Matches	409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	12	GCCTCCAGCATGGCGCTTCGCTCTCCACGACACCGGTCTACTGCTTGAGTGTAT 71
Db	8451	GCCTCCAGCATGGCGCTTCGCTCTCTCCACGACACCGGTCTACTGCTTGAGTGTAT 8392
Qy	72	ACCAGGTAAATTAAGCCCAACACTCAGCTGGGGGAGCATCAAGCGCAGAGGCCATG 131
Db	8391	ACCAGGTAAATTAAGCCCAACACTCAGCTGGGGGAGCATCAAGCGCAGAGGCCATG 8332
Qy	132	TTGGCCAAAGAGCTCCAGGCTCTATGGGTATATGAACCTTGAGGGCGGTGTAAACCA 191
Db	8331	TTGGCCAAAGAGCTCCAGGCTCTATGGGTATATGAACCTTGAGGGCGGTGTAAACCA 8272
Qy	192	AGAAAGCATGAACCGGATGGGGGAGTGAATCCCGAGGGGGCGGTGTGAACCAAGAGC 251
Db	8271	AGAAAGCATGAACCGGATGGGGGAGTGAATCCCGAGGGGGCGGTGTGAACCAAGAGC 8212
Qy	252	CTCGGACAGACCGCCAGCCCGCGGAGGAGCCGCGCCAGCGCCATGAGGGGTCCGC 311
Db	8211	CTCGGACAGACCGCCAGCCCGCGGAGGAGCCGCGCCAGCGCCATGAGGGGTCCGC 8152
Qy	312	AGCCCGAATCATGACGCCACGAATCTGGCGGCACGAAGTTTCACTGTTCAGGTGAGG 371
Db	8151	AGCCCGAATCATGACGCCACGAATCTGGCGGCACGAAGTTTCACTGTTCAGGTGAGG 8092
Qy	372	AGCTGAAAGTGTTCGACACACTCAATACCTGATGAGGCCACAG 420
Db	8091	AGCTGAAAGTGTTCGACACACTCAATACCTGATGAGGCCACAG 8043
RESULT 5		
AC005023/c	170967 bp	DNA linear PRI 21-DEC-1999
LOCUS	AC005023	
DEFINITION	Homo sapiens BAC clone GSI-42113 from Xq25-q26, complete sequence.	
ACCESSION	AC005023	
VERSION	AC005023.1 GI:3900847	
KEYWORDS	HTG.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS	1 (bases 1 to 170967)	
TITLE	Leonard, S., Graves, T. and Cotman, M.	
REFERENCE	The sequence of Homo sapiens BAC clone GSI-42113	
JOURNAL	Unpublished	
AUTHORS	2 (bases 1 to 170967)	
TITLE	Waterston, R.H.	
REFERENCE	Direct Submission	
JOURNAL	Submitted (12-UTN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	
AUTHORS	3 (bases 1 to 170967)	
TITLE	Waterston, R.	
REFERENCE	Direct Submission	
JOURNAL	Submitted (21-NOV-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	
AUTHORS	4 (bases 1 to 170967)	
TITLE	Waterston, R.	
REFERENCE	Direct Submission	
JOURNAL	Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	
COMMENT	On Nov 21, 1998 this sequence version replaced gi:3212968.	

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_GS421103

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This clone was provided for sequencing by Buddy Brownstein in the Center for Genetics in Medicine and by John D. McPherson in the Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).
Cell line: lymphoblastoid
Haplotypes: two
VECTOR: pBelobAC
Selection: chloramphenicol
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP3-327A19, 200 bp overlap. The actual start of this clone is at base position 88746 of RP3-327A19; actual end is at 170967 of GS1-421I3.
Location/Qualifiers

FEATURES

source
1..170967
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="XG25-q26"
/clone_1fb="GSBAC1"
1..37
/rpt_family="Alu"
38..258
/rpt_family="Alu"
273..568
/rpt_family="Alu"
1789..1835
/rpt_family="polypurine"
2046..2191
/rpt_family="L1"
2192..2218
/rpt_family="(CAA)n"
2222..2517
/rpt_family="Alu"
2518..2816
/rpt_family="L1"
2817..2889
/rpt_family="Retroviral"
2892..3011
/rpt_family="(TA)n"
3063..3084
/rpt_family="AT-rich"
3085..3493
/rpt_family="Retroviral"
3509..4189

repeat_region /rpt_family="L1" 4190..4481
repeat_region /rpt_family="Alu" 4482..4662
repeat_region /rpt_family="L1" 4663..4960
repeat_region /rpt_family="Alu" 4961..5869
repeat_region /rpt_family="L1" 5870..5890
repeat_region /rpt_family="(TAAAA)n" 5891..6624
repeat_region /rpt_family="L1" 6625..6960
repeat_region /rpt_family="Retroviral" 6961..7095
repeat_region /rpt_family="L1" 7096..7274
repeat_region /rpt_family="Alu" 7275..7509
repeat_region /rpt_family="L1" 7510..7791
repeat_region /rpt_family="Alu" 7792..8370
repeat_region /rpt_family="L1" 8371..8669
repeat_region /rpt_family="Alu" 8680..9119
repeat_region /rpt_family="Retroviral" 9120..9194
repeat_region /rpt_family="MER4-group" 9195..9431
repeat_region /rpt_family="Retroviral" 9505..9585
repeat_region /rpt_family="Alu" 9669..9699
repeat_region /rpt_family="L1" 9701..9721
repeat_region /rpt_family="(TAAA)n" 9722..9996
repeat_region /rpt_family="Alu" 9997..10338
repeat_region /rpt_family="L1" 10368..10544
repeat_region /rpt_family="Alu" 10562..10884
repeat_region /rpt_family="L1" 10885..11177
repeat_region /rpt_family="Alu" 11178..11207
repeat_region /rpt_family="L1" 11208..11512
repeat_region /rpt_family="Alu" 11513..12024
repeat_region /rpt_family="L1" 12025..12283
repeat_region /rpt_family="Alu" 12284..12603
repeat_region /rpt_family="L1" 12604..12886
repeat_region /rpt_family="Alu" 12887..12974
repeat_region /rpt_family="L1" 12975..13118
repeat_region /rpt_family="Alu" 13121..14063
repeat_region /rpt_family="L1" 14122..16240
repeat_region /rpt_family="L1" 16241..16536
repeat_region /rpt_family="Alu" 16537..19982
repeat_region /rpt_family="L1"

```

repeat_region      20009..20210
                    /rpt_family="Alu"
repeat_region      20211..20511
                    /rpt_family="Alu"
repeat_region      20512..20617
                    /rpt_family="Alu"
repeat_region      20627..21318
                    /rpt_family="Alu"
repeat_region      21320..21363
                    /rpt_family="AT-rich"
repeat_region      22857..22971
                    /rpt_family="Alu"
repeat_region      23001..23031
                    /rpt_family="GC-rich"
repeat_region      23085..23630
                    /rpt_family="MER4-group"
repeat_region      24345..24373
                    /rpt_family="AT-rich"
repeat_region      24431..24549
                    /rpt_family="Alu"
repeat_region      24762..25060
                    /rpt_family="Alu"
repeat_region      25786..26085
                    /rpt_family="Alu"

```

Query Match 70.9%; Score 409; DB 9; Length 170967;
 Best Local Similarity 100.0%; Pred. No. 1.4e-77;
 Matches 409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 12 GCGCTCCAGCCATGCGCGCTGCTGTCACAGACACCGTGTCTACTGCTGAGTGAT 71
DB 168450 GCGCTCCAGCCATGCGCGCTGCTGTCACAGACACCGTGTCTACTGCTGAGTGAT 168391
QY 72 ACCAGTAAATAAATAGAGCCCACTCCTGAGTGGGAGACATCAAGCCGAAAGGCCATG 131
DB 168390 ACCAGTAAATAAATAGAGCCCACTCCTGAGTGGGAGACATCAAGCCGAAAGGCCATG 168331
QY 132 TTGGCCAGAGAGCTCCAGGCTCATGAGGTAATATGAACCTGAGGGCGGTGGAACCAAG 191
DB 168330 TTGGCCAGAGAGCTCCAGGCTCATGAGGTAATATGAACCTGAGGGCGGTGGAACCAAG 168271
QY 192 AGAAGCGCATGAACCGCGATGCGCGCATGATCCCGAGGGCGCGGTGGAACCAAGAGC 251
DB 168270 AGAAGCGCATGAACCGCGATGCGCGCATGATCCCGAGGGCGCGGTGGAACCAAGAGC 168211
QY 252 CTCGGCAGACGCGCAGCCCGCGGAGAGCGCGGCCCAAGGGCGCATGAGGGTCCGC 311
DB 168210 CTCGGCAGACGCGCAGCCCGCGGAGAGCGCGGCCCAAGGGCGCATGAGGGTCCGC 168151
QY 312 AGCCGAGAAATGACGACGCAACTCGGCGGACGAAAGTTCAAGCTGTGCAAGTGGAGG 371
DB 168150 AGCCGAGAAATGACGACGCAACTCGGCGGACGAAAGTTCAAGCTGTGCAAGTGGAGG 168091
QY 372 AGCTGGAAGTGTTCGACACACTCAATACCTGATGTGCCCAAG 420
DB 168090 AGCTGGAAGTGTTCGACACACTCAATACCTGATGTGCCCAAG 168042

```

RESULT 6
 AC145687 139483 bp DNA linear HTG 23-AUG-2003
 LOCUS AC145687/c Pan troglodytes chromosome UNK clone CH251-281C1, WORKING DRAFT
 DEFINITION
 AC145687 3 unordered pieces.
 AC145687
 AC145687.3 GI:34147034
 VERSION HTG, HTGS_PHASE1, HTGS_DRAFT, HTGS_ACTIVEFIN.
 KEYWORDS Pan troglodytes (chimpanzee)
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 1 (bases 1 to 139483)
 REFERENCE
 AUTHORS
 TITLE
 The sequence of Pan troglodytes clone

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished
 2 (bases 1 to 139483)
 Wilson, R.K.
 Direct Submission
 Submitted (19-JUL-2003) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA
 3 (bases 1 to 139483)
 Wilson, R.K.
 Direct Submission
 Submitted (23-AUG-2003) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA
 On Aug 23, 2003 this sequence version replaced gi:3387229.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu
 Contact: submissions@watson.wustl.edu
 Project information
 Center project name: C_AA0281C01

----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: Plasmid; 100%
 Chemistry: Dye-primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 137838 bases at least Q40
 Consensus quality: 138075 bases at least Q30
 Consensus quality: 138228 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 configs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the configs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1266: contig of 1266 bp in length
 * 1267 1366: gap of unknown length
 * 1367 51720: contig of 50354 bp in length
 * 51721 51820: gap of unknown length
 * 51821 139483: contig of 87663 bp in length.

FEATURES

source
 1..139483
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /chromosome="UNK"
 /clone="CH251-281C1"
 1..1266
 /note="assembly_name:Contig5"
 1367..51720
 /note="assembly_name:Contig17"
 51821..139483
 /note="assembly_name:Contig18"

ORIGIN

Query Match 69.2%; Score 399.4; DB 2; Length 139483;
 Best Local Similarity 98.5%; Pred. No. 1.6e-75;
 Matches 403; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

QY 12 GCGCTCCAGCCATGCGCGCTGCTGTCACAGACACCGTGTCTACTGCTGAGTGAT 71
DB 39897 GCGCTCCAGCCATGCGCGCTGCTGTCACAGACACCGTGTCTACTGCTGAGTGAT 39838
QY 72 ACCAGTAAATAAATAGAGCCCACTCCTGAGTGGGAGACATCAAGCCGAAAGGCCATG 131
DB 39837 ACCAGTAAATAAATAGAGCCCACTCCTGAGTGGGAGACATCAAGCCGAAAGGCCATG 39778
QY 132 TTGGCCAGAGAGCTCCAGGCTCATGAGGTAATATGAACCTGAGGGCGGTGGAACCAAG 191

```

```

Db      39777 TTGGCCAGAGAGCTCCAGGCTTCATGGTAATATGAAACCTTGAGGGCGGTGTGAACACAG 39718
Qy      192 AGAAGCGCATGAAACCGCATATGGCGCATATGCCCGAGGGCGCGGTGAAACACAGAGC 251
Db      39717 AGAACGGCATGAACCGCATATGGCGCATATGCCCGAGGGCGCGGTGAAACACAGAGC 39658
Qy      252 CTCGGCAGCAGCCGCGAGCCCGCGGAGAGCGCGGCCCGCATGAGAGGTCCGC 311
Db      39657 CTAGCGCAGCAGCCGCGAGCCCGCGGAGAGCGCGGCCCGCATGAGAGGTCCGC 39598
Qy      312 AGCCCGAAGAACATGACGCGCAGCACTCGGGCAGCAAGTTGAGCTGTTGACAGTGAAG 371
Db      39597 AGCCCGAAGAACATGACGCGCAGCACTCGGGCAGCAAGTTGAGCTGTTGACAGTGAAG 39538
Qy      372 AGCTGGAAGTGTTCCTCGACACACTCAATACCTGATGTGCCACAG 420
Db      39537 AGCTGGAAGTGTTCCTCGACACACTCAATACCTGATGTGCCACAG 39489

RESULT 7
AC146356/c
LOCUS      204520 bp      DNA      linear      HTG 27-SEP-2003
DEFINITION Pongo pygmaeus clone CH253-378A10, WORKING DRAFT SEQUENCE, 6
ACCESSION AC146356
VERSION    AC146356.2 GI:36430306
KEYWORDS   HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE     Pongo pygmaeus (orangutan)
ORGANISM   Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pongo.
            1 (bases 1 to 204520)
REFERENCE   1 (bases 1 to 204520)
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pongo.
            Ayele, D.C., Rozen, S., Salenz, J.R., Skaletsky, H., Antonellis, A.,
            Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W.,
            Bouffard, G.G., Brinkley, C., Brooks, S., Cariaga, K., Chu, G.,
            Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X., Gupta, J.,
            Haghighi, P., Hansen, N., Ho, S.-L., Hu, P., Hulse, B.,
            Idol, J.R., Karlins, E., Kwong, P., Laric, P., Lee-Hin, S.-Q.,
            Legaspi, R., Maduro, Q.L., Maduro, V.B., Marquiles, E.H., Mastello, C.,
            Maskeri, B., McDowell, J., Paguirigan, C., Pearson, R., Portnoy, M.E.,
            Praead, A., Reddik-Dugue, N., Schandler, K., Schueler, M.G., Shah, K.,
            Sison, C., Stantropop, S., Thomas, J.W., Thomas, P.J., Tsipouri, V.,
            Vogt, J.L., Wetherly, K.D., Young, A. and Green, E.D.
            NISC Comparative Sequencing Initiative
            Unpublished
            2 (bases 1 to 204520)
TITLE      Direct Submission
JOURNAL    Submitted (13-AUG-2003) NIH Intramural Sequencing Center, 8717
AUTHORS    Grovumot Circle, Galtersburg, MD 20877, USA
JOURNAL    3 (bases 1 to 204520)
REFERENCE   Green, E.D.
TITLE      Direct Submission
JOURNAL    Submitted (27-SEP-2003) NIH Intramural Sequencing Center, 8717
AUTHORS    Grovumot Circle, Galtersburg, MD 20877, USA
JOURNAL    On Sep 27, 2003 this sequence version replaced gi:33620791.
COMMENT    ----- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_zoo@nhgri.nih.gov
            ----- Project Information
            Center project name: erl
            Center clone name: 378A10

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig, has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average

coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 203385 bases at least Q40
 Consensus quality: 203760 bases at least Q30
 Consensus quality: 203958 bases at least Q20
 Insert size: 211000; agarose-fp
 Insert size: 204020; sum-of-contigs
 Quality coverage: 9.32x in Q20 bases; agarose-fp
 Quality coverage: 9.64x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

```

* 1 10013: contig of 10013 bp in length
* 10014 10113: gap of unknown length
* 10114 14402: contig of 4289 bp in length
* 14403 14502: gap of unknown length
* 14503 79161: gap of 64559 bp in length
* 79162 79261: gap of unknown length
* 79262 156374: contig of 77113 bp in length
* 156375 156474: gap of unknown length
* 156475 188945: contig of 32471 bp in length
* 188946 189046: gap of unknown length
* 189046 204520: contig of 15475 bp in length.
Location/Qualifiers
1. 204520
/organism="Pongo pygmaeus"
/mol_type="genomic DNA"
/db_xref="taxon:9600"
/clone="CH253-378A10"
/clone_1fb="CH253"
1. 10013
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
10114..14402
/note="assembly_fragment"
14503..79161
/note="assembly_fragment"
79262..156374
/note="assembly_fragment"
156475..188945
/note="assembly_fragment"
189046..204520
/note="assembly_fragment"
missing T7 clone end on 3' end of insert"

```

```

misc_feature
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
10114..14402
/note="assembly_fragment"
14503..79161
/note="assembly_fragment"
79262..156374
/note="assembly_fragment"
156475..188945
/note="assembly_fragment"
189046..204520
/note="assembly_fragment"
missing T7 clone end on 3' end of insert"

```

ORIGIN

```

Query Match      66.4%; Score 383.4; DB 2; Length 204520;
Best Local Similarity 96.1%; Pred. No. 4.2e-72;
Matches 393; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy      12 GCGCTCCAGCAATGGCGGCTGCTGTCACAGACCGGTGTTCTACCTGAGTGAT 71
Db      164300 GCGCTCCAGCAATGGCGGCTGCTGTCACAGACCGGTGTTCTACCTGAGTGAT 164241

Qy      72 ACCAGGTAATAATTAAGCCCAACACTCAGCTGGGGCAGCATCAAGCGAGAGCCATG 131
Db      164240 ACCAGGTAATAATTAAGCCCAACACTCAGCTGGGGCAGCATCAAGCGAGAGCCATG 164181

Qy      132 TTGGCCAGAGAGCTCCAGGCTTCATGGTAATATGAAACCTTGAGGGCGGTGTGAACACAG 191

```


QY	252	CTGGGACGACGACCGCAGACCCCGCCGGAGAGAGCGCGCCAGAGCGGACCATGAGAGATCCGC	311
Db	76404	CTAGGACGCTGCGACAGACCC---CTGAGAGAGCCCGGCCAGCGGCGCATGAGAGATCCGC	76466
QY	312	AGCCCCGAGAACATGACAGCCACGAACTTGCGGCGCAGGAATTCACGCTGTTCAGGTGAGAG	371
Db	76461	AGCCCCGAGAACATGACAGCCACGAAATTTGGGGGCAAGAAATTCACGCGGTTCCAGGTCCAGG	76520
QY	372	AGCTGGAAAGTGTGTTTTCCGACACACTCAATATCCCTGATGTGCCCCACAAAG	420
Db	76521	AGCTGGAAAGTGTGTTCTCAACGACATCAATATCCCTGATGTGCCCCACAAAG	76569
RESULT 9			
LOCUS	HS421131	1205 bp	mRNA linear PRI 11-APR-2001
DEFINITION	Novel human mRNA from X chromosome, contains homeobox.		
ACCESSION	AL590524		
VERSION	AL590524.1	GI:13620481	
KEYWORDS	.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Hominidae; Homo.		
	1 (bases 1 to 1205)		
REFERENCE	Howell,G.R., Huckle,E. and Ross,M.T.		
AUTHORS	Direct Submission		
TITLE	Submitted (09-APR-2001) E-mail contact: humquery@sanger.ac.uk		
JOURNAL	This cDNA sequence was assembled from public domain ESTs and single		
COMMENT			

FEATURES	
source	location/Qualifiers
	1..1205
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/chromosome="X"
	/map="Xq23-Xq24"
	1..165
exon	/number=1
	87..953
CDS	/codon_start=1
	/product="hypothetical protein"
	/protein_id="CAC36517.1"
	/db_xref="GI:13620482"
	/db_xref="GOA:Q9BR00"
	/db_xref="SPTRMBL:Q9BR00"
	/translation="MEPPDCSQYMTSLSPAVDKEKLQDMANMVLSTEEVEEEREE DADEPEQGTAAEGKLSAGAGGEEKDGGAGVPGILMGNGLEGTSGDG NVESDSEKERQQOYSRPGAVNGLERGNAGQNVHYAFPILOJQLCELCIFORPFS EFLRRRLAKSMNTVELAVQIWFENRRAKRRHQALMARNNLPMAVGQPVVTAAER ITVALFTSGMRDYFWDSSHSSSLCFMPPPPSPSLPLMLLPMPAPGAERGPPPP FVIAVSFPFPNV"
	166..577
exon	/number=2
	578..623
exon	/number=3
	583..859
misc_feature	/note="matches EST R69179 from clone IMAGE:141589"
	624..1205
exon	/number=4
	complement(join(730..956,952..1203))
misc_feature	/note="matches EST A1954612 from clone IMAGE:2473139"
	complement(739..1204)
misc_feature	/note="matches EST BB046930 from clone IMAGE:2916991"
	complement(788..1204)
misc_feature	/note="matches EST A1657125 from clone IMAGE:2244151"
	complement(798..1205)

misc_feature	/note="matches EST AI1209091 from clone IMAGE:1838548"
	complement(811...1200)
ORIGIN	/note="matches EST AA933590 from clone IMAGE:1551106"
Query Match	18.5%; Score 106.8; DB 9; Length 1205;
Best Local Similarity	60.7%; Pred. No. 1,16-12;
Matches 193; Conservative	0; Mismatches 122; Indels 3; Gaps 1;
Dy	203 AACCGCATGAGCGGCATGATATCCCGAAGGCGCGCGTGTGAAACCAAGACCTCGCAGCAG 262
Dy	363 AACCTCGAAGGCACCAACGCGAGGATGCAACCTTAGAGCAACAGACCAAGAGGAGAG 422
Dy	263 CCGCAGCCCCCGCCGAGAGAGCCGCGCCAGCGGCCCATGAGGGTCCGACGCCGAGAAC 322
Dy	423 GAACCTTGGGAGCAGATTTGCGGCCCAACAGGGCCCGCTCGGGGGGTGTGAACTTCGCAAC 482
Dy	323 ATGCAAGCACCAACCTCGGCGCACAAATTACGCTGTTCAGGTGAGTGTGAGAGACTGTGAAAGT 382
Dy	483 GCGGAGAGAGCCCAACGTTCAAGC---CTTCAACCCCATTTGCAGCTGTGAGAGACTGTGAGTGC 539
Dy	383 GTTTTCCGACACACTCAATACCTTGATGTGTGCCCAAGAAAGGAACTTGTCCGAAACTTA 442
Dy	540 ATTTTCCACGCGAGCAGTTCTCCAGTAGTCTTCGCAAGAGAGAGCTGTGCAAGAGCAATG 599
Dy	443 GGTGTGACTGTAGAACAAAGTGTGGGTTTGTGTTTAAGATTAAGAGGCGCCAGATGTAGCGCA 502
Dy	600 AATGTGACTGAACTCGCAGTGCAGATTTGGTTTGAGATTAAGAGCCAAATGTGAGGAGAGA 659
Dy	503 CATCAGAGAGATTTAATG 520
Dy	660 CATCAGAGGCGATTTAATG 677

RESULT	10
LOCUS	AKO58125
DEFINITION	AKO58125 1194 bp mRNA linear PRI 12-SEP-2003
ACCESSION	AKO58125
VERSION	AKO58125.1 GI:16554167
KEYWORDS	oligo capping; fls (full insert sequence).
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S., Ono,Y., Hottta,T., Hirooka,S., Murakawa,K., Takiguchi,S., Kusano,U., Watanabe,M., Fujimori,K., Tanai,H., Ihida,M., Yamasaki,H., Chiba,Y., Suzuki,Y., Hate,H., Nakagawa,K., Mizuno,S., Morinaga,M., Kawamura,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Nishikawa,T., Sugiyama,A., Kawakami,B., Nagai,K., Isogai,T. and Sugano,S. NEDO human cDNA sequencing project
TITLE	Unpublished
JOURNAL	2 (bases 1 to 1194)
REFERENCE	Sugano,S. and Suzuki,Y.
AUTHORS	Direct Submission
TITLE	Submitted (24-OCT-2001) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:tlcdna@ims.u-tokyo.ac.jp, Tel :81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction and 5'-end one pass sequencing: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; 3'-end one pass sequencing: RAB; clone selection for full insert sequencing: RAB and Helix Research Institute.
FEATURES	Location/Qualifiers 1..1194 /organism="Homo sapiens"
SOURCE	

CDS

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TST02591"
/tissue_type="testis"
/clone_lib="TST"
/notes="Cloning vector: pME18SFL3"
84..950
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB71675.1"
/db_xref="GI:16554168"
/translation="MEPPDQCSQYMTSLSPAVDDKEKLDNNMNVLSLTERVEKEE
DAQPEPQGTAGEKLSKAGQGEKQGEKQGGAGVGHLMEDLIGTSGSDG
NVEDDSEKQYRPOGAVGLPEGNACQPVNVAFTPIQLDELRIPIREDFPS
EFRRRLARSMNVTELAVOIWEENRBAKRHRORALMARNMLPFMAVQCPVNTAAE
ITAPLFISGMDDYFWDHSHSSSLCFPMPPPPPLPLPLMLPMPMPAGQAEFPFP
FVIYPSFTFPNV"

```

ORIGIN

Query Match 18.3%; Score 105.8; DB 9; Length 1194;
 Best Local Similarity 60.6%; Pred. No. 1.8e-12;
 Matches 192; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

```

204 ACCGGGATGCGCGCATGATCCCGGAGGCGGCGGTGGAACACGAGCCTCGGACAGC 263
361 ACCTGAGAGGACACGCGGACGATGCAAGCTTGAGACAGCAGCAGACGAGAGG 420
264 CCGACGCCCCCGCGGAGAGCGCGGCGGCGGACATGAGAGGTCCGACGCCGAGACA 323
421 AACCTGGGAGAGCATTCGCGCCACAGGCGCGCGTCCGCGGCGTGGAGCCTGGCAACG 480
324 TGCAGCCACGAACTGGGCGGACGAGTTCAGCTGTGAGGTGAGAGGAGCTGGAAAGTG 383
481 CGCAGCAGCCCAAGCTCCAGCG---CTTACCCCATATGAGCTGCGAGGAGCTGAGCGCA 537
384 TTTTCCGACACATGATCCCTGATGTGCCCAAGAAGGAACTTGGCGGAAACTTAG 443
538 TTTTCAACGCGAGAGAGTCCCGAGTGTCTGCGAAGAGAGGCTGCAAGAGATGA 597
444 GTGTGACGTGAAGACAAAGTGGCGGTTGGTTTAAGATTAAGAGCCAGATGTAGCGAC 503
598 ATGTGACTGTAAGCTCCAGAGTGAAGTTGGTTGAGATGAAGAGCAATGAGAGAGAC 657
504 ATCAGAGAGCAATTAATG 520
658 ATCAGAGGCGCATTAATG 674

```

RESULT 11

BC021719 1219 bp mRNA linear PRI 08-OCT-2003
 LOCUS Homo sapiens PEPP subfamily gene 2, mRNA (cDNA clone MGC:32918
 DEFINITION IMAGE:5269333), complete cde.

ACCESSION BC021719
 VERSION BC021719.2 GI:34194251
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1219)
 AUTHORS Struhsberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldi,M.F., Casavant,T.L., Scheetz,T.R., Brownstein,M.J., Ueda,T.B., Toshiyuki,S., Carinuci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,

Faney,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Buterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalins,D.E., Scherich,A., Schein,J.E., Jones,S.J., and Marra,M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE 22388257
 PUBMED 12477932

REFERENCE 2 (bases 1 to 1219)

Struhsberg,R.

Direct Submission

Submitted (14-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 On Aug 25, 2003 this sequence version replaced gi:18204436.

COMMENT REMARK
 CONTACT: MGC help desk
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>

Contact: amadan@systemsbio.org
 Anup Madan, Jessica Faney, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IBK Plate: 47 Row: 0 Column: 18
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 14249118.

FEATURES
 source Location/Qualifiers
 1..1219
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:32918 IMAGE:5269333"
 /tissue_type="testis"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Vector: pBluescript"

1..1219
 /gene="PEPP-2"
 /db_xref="LocusID:84528"
 /db_xref="MIM:300448"
 86..952
 /codon_start=1
 /product="homeobox protein from AL590526"
 /protein_id="AAH21719.1"
 /db_xref="GI:18204437"
 /db_xref="LocusID:84528"
 /translation="MEPPDQCSQYMTSLSPAVDDKEKLDNNMNVLSLTERVEKEE
 DAQPEPQGTAGEKLSKAGQGEKQGEKQGGAGVGHLMEDLIGTSGSDG
 NVEDDSEKQYRPOGAVGLPEGNACQPVNVAFTPIQLDELRIPIREDFPS
 EFRRRLARSMNVTELAVOIWEENRBAKRHRORALMARNMLPFMAVQCPVNTAAE
 ITAPLFISGMDDYFWDHSHSSSLCFPMPPPPPLPLPLMLPMPMPAGQAEFPFP
 FVIYPSFTFPNV"

gene
 CDS

misc_feature
 /note="homeobox; Region: Homeobox domain"
 /db_xref="CDD:pfam00046"

ORIGIN

Query Match 18.3%; Score 105.8; DB 9; Length 1219;
 Best Local Similarity 60.6%; Pred. No. 1.8e-12;
 Matches 192; Conservative 0; Mismatches 122; Indels 3; Gaps 1;
 204 ACCGGGATGCGCGCATGATCCCGGAGGCGGCGGTGGAACACGAGCCTCGGACAGC 263

```

Db      363  ACCCTCAGGGGCAACCGAGGAGATGGCAACCTTAGAGACAGCGACCGAGCGAAGAG 422
Qy      264  CGCAGCCCCCGCGGAGAGACCGGCCAGGCGGCATGAGAGGATCCGACGCCGAGAAACA 323
Db      423  AACCTGGGAGAGAGATTTGCGCCGCCACAGGGGCCCTCGGGGGGCTGGAGCCTGGCAAG 482
Qy      324  TGCAGCACAAGACTCGGCGCAGCAAGTTCAAGCTGTTCAGAGTGAAGAGCTTGAAAGTG 383
Db      483  CGCAGAGCGCCCAACGTCCACGC---CTTCACCCCATTTGACGTGACAGAGCTGAGCGCA 539
Qy      384  TTTTCGACACACTCAATACCTGATGTGCCCAAGAGAGAACTTGCAGAAACTTAG 443
Db      540  TTTTCCAGCGAGCAGCTTCCCAAGTAGATTCTGCGAAGAGAGCTGGCAGAAAGCATGA 599
Qy      444  GTGTGACTGAAGACAAGTGCAGGTTTGGTTTAAGATAAAAGGCGCAGATGAGCGAC 503
Db      600  ATGTGACTGAACCTCGCAGTGCAGATTGGTTTGAAGATGAGAGAGCCAAATGAGAGAGAC 659
Qy      504  ATCAGAGAGATTATG 520
Db      660  ATCAGAGGCGATTATG 676

```

```

RESULT 12
AF317219      1274 bp      mRNA      linear      PRI 13-SEP-2001
LOCUS        Homo sapiens homeobox protein (THG1) mRNA, complete cds.
DEFINITION   AF317219
ACCESSION    AF317219.1 GI:15592921
VERSION      AF317219.1 GI:15592921
KEYWORDS
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 1274)
AUTHORS      Yin,Y., Jin,Y. and Levine,A.J.
TITLE        Molecular cloning and characterization of a novel testis homeobox
              gene, THG1, as a potential testicular tumor suppressor gene
              Unpublished
              2 (bases 1 to 1274)
              Submitted (26-OCT-2000) Radiation Oncology, Columbia University,
              630 West, 168th Street, New York, NY 10032, USA
              Location/Qualifiers
                1..1274
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /chromosome="X"
                /tissue_type="testis"
                1..1274
                /gene="THG1"
                /gene="THG1"
                131..997
                /gene="THG1"
                /codon_start=1
                /product="homeobox protein"
                /protein_id="AA102160.1"
                /db_xref="GI:15592922"
                /translation="MEPPDCSQYMTSLSPAVDDEKELQDNAMVLSLTYEKREE
                DADPEPQGTAAAGEKLSKSAAGGEEKDGGGAGVPGHLWEGDLEGGSGDG
                NWEDSDQSEKEPQOYSRPGAVGLEPGNAQPNVHAFTPLQLEIERIFQRECPDS
                EFLRRRLASKNVTELAIVOIFENRRAKRRHORALMARMLPFAVVGQPVNTAAEA
                ITAPLFISGMRDYFMDHSHSSSLCPMPFPFPPSLPLPLMLPMPRPAAGQAFGFPF
                FVIVPSFTFPNV"

```

ORIGIN

```

Query Match      18.3%; Score 105.8; DB 9; Length 1274;
Best Local Similarity 60.6%; Pred. No. 1.8e-12;
Matches 192; Conservative 0; Mismatches 122; Indels 3; Gaps 1;
Qy      204  ACCGCATGCGCGCATGATCCCGAGGGCGCGGTGGAACAAGAGCCTCGCAGACGAC 263

```

```

Db      408  ACCCTCAGGGGCAACCGGAGAGATGGCAAGTTGAGACAGCGACCGAGCGAAGAG 467
Qy      264  CGCAGCCCCCGCGGAGAGACCGGCCAGGCGGCATGAGAGGATCCGACGCCGAGAAACA 323
Db      468  AACCTGGGAGAGAGATTTGCGCCGCCACAGGGGCCCTCGGGGGGCTGGAGCCTGGCAAG 527
Qy      324  TGCAGCACAAGACTCGGCGCAGCAAGTTCAAGCTGTTCAGAGTGAAGAGCTTGAAAGTG 383
Db      528  CGCAGAGCGCCCAACGTCCACGC---CTTCACCCCATTTGACGTGACAGAGCTGAGCGCA 584
Qy      384  TTTTCGACACACTCAATACCTGATGTGCCCAAGAGAGAACTTGCAGAAACTTAG 443
Db      585  TTTTCCAGCGAGCAGCTTCCCAAGTAGATTCTGCGAAGAGAGCTGGCAAGAGCATGA 644
Qy      444  GTGTGACTGAAGACAAGTGCAGGTTTGGTTTAAGATAAAAGGCGCAGATGAGCGAC 503
Db      645  ATGTGACTGAACCTCGCAGTGCAGATTGGTTTGAAGATGAGAGAGCCAAATGAGAGAGAC 704
Qy      504  ATCAGAGAGATTATG 520
Db      705  ATCAGAGGCGATTATG 721

```

```

RESULT 13
HS525N141      1297 bp      mRNA      linear      PRI 11-APR-2001
LOCUS        Novel human mRNA from X chromosome, contains homeobox.
DEFINITION   HS525N141
ACCESSION    AL590526
VERSION      AL590526.1 GI:13620840
KEYWORDS
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 1297)
AUTHORS      Howell,G.R., Huckle,E. and Ross,M.T.
TITLE        Direct Submision
              Submitted (09-APR-2001) E-mail contact: humquery@sanger.ac.uk
              This cDNA sequence was assembled from public domain ESTs and single
              pass sequencing reads from expressed DNA templates, aligned to the
              genomic DNA sequence from the bacterial clone 525N14 (AC002086)
              The EST sequences listed match this sequence with an identity of at
              least 95% between the coordinates shown.
              Further information can be found at
              http://www.sanger.ac.uk/HGP/Chx/ Sanger Centre name :
              dJ525N14.CX.1.
              Location/Qualifiers
                1..1297
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /chromosome="X"
                /map="Xq23-Xq24"
                1..269
                /number=1
                191..1057
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAC36519.1"
                /db_xref="GI:13620841"
                /db_xref="GOA:Q9BOY4"
                /translation="MEPPDCSQYMTSLSPAVDDEKELQDNAMVLSLTYEKREEE
                DADPEPQGTAAAGEKLSKSAAGGEEKDGGGAGVPGHLWEGDLEGGSGDG
                NWEDSDQSEKEPQOYSRPGAVGLEPGNAQPNVHAFTPLQLEIERIFQRECPDS
                EFLRRRLASKNVTELAIVOIFENRRAKRRHORALMARMLPFAVVGQPVNTAAEA
                ITAPLFISGMRDYFMDHSHSSSLCPMPFPFPPSLPLPLMLPMPRPAAGQAFGFPF
                FVIVPSFTFPNV"

```

```

exon      270..681
exon      682..727
exon      /number=2
exon      /number=3

```

```
misc_feature      687..963
                  /note="matches EST R69179 from clone IMAGE:141589"
exon              728..1297
                  /number=4
misc_feature      complement(join(834..1060,1056..1297))
                  /note="matches EST A1954612 from clone IMAGE:2471339"
misc_feature      complement(843..1297)
                  /note="matches EST BE046930 from clone IMAGE:2916991"
misc_feature      complement(892..1297)
                  /note="matches EST A1857125 from clone IMAGE:2244151"
misc_feature      complement(902..1297)
                  /note="matches EST A1209091 from clone IMAGE:1838548"
misc_feature      complement(915..1297)
                  /note="matches EST A933590 from clone IMAGE:1551106"
misc_feature      complement(1066..1296)
                  /note="matches EST R69071 from clone IMAGE:141589"
ORIGIN
Query Match      18.3%; Score 105.8; DB 9; Length 1297;
Best Local Similarity 60.6%; Pred. No. 1.8e-12;
Matches 192; Conservative 0; Mismatches 122; Indels 3; Gaps 1;
QY 204 ACCGCGATGCGCGCATGATCCCGAGCGCGCGGTGAAAACGAGAGCCTCGGACAGC 263
DB 468 ACCTCGAGGGCACACGCGGACGATGGCAACGTTGAGGACAGCGACGAGGAGAGG 527
QY 264 CCCAGCCCCCGCGAGAGCCGCGCCCATGAGGGTCCGCGACCCGAGAAACA 323
DB 528 AACCTGGGAGAGATGTTGGCGCCGACAGGCGCCCTCGGGGGCTGGAGCCTGGACAG 587
QY 324 TCGAGCCAGAACTCGGCGCAGAACTGACGCTGTTGAGGTGAGAGAGCTGGAAGTG 383
DB 588 CGCAGAGAGCCCAACGTCACAGC---CTTACCCCATTTGACGTCGAGAGCTGAGAGCGCA 644
QY 384 TTTTCGACACACTCAATACCTGATGTGCCCAAGAAAGAACTTGCCGAATCTTGA 443
DB 645 TTTTCCACGCGAGCGAGTCCCGAGTGAAGTTCCTGGAGAGGCTGGCAAGAGCATGA 704
QY 444 GGTGACCTAAGACAAAGTCCGGTGTGTTTAAAGAAAGGCGCAATGATGAGCGAC 503
DB 705 ATGTGACTAAGCTCGAGTGCAGATTGTTGATGAATGAAAGCAATGAGAGAGAC 764
QY 504 ATCAGAGAGATTATG 520
DB 765 ATCAGAGGCGCATTAATG 781
RESULT 14
AY114148          1451 bp      mRNA      linear      PRI 29-JUN-2002
LOCUS             Homo sapiens ESX1L (ESX1L) mRNA, complete cde.
ACCESSION         AY114148
VERSION           AY114148.1 GI:21629645
KEYWORDS
SOURCE            Homo sapiens (human)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE         1 (bases 1 to 1451)
AUTHORS           Fohn,L.E. and Behringer,R.R.
TITLE             ESX1L, a novel X chromosome-linked human homeobox gene expressed in
                  the placenta and testis
JOURNAL           Genomics 74 (1), 105-108 (2001)
MEDLINE           21269184
PUBMED            11374906
REFERENCE         2 (bases 1 to 1451)
AUTHORS           Fohn,L.E. and Behringer,R.R.
TITLE             Direct Submision
JOURNAL           Submitted (17-MAY-2002) Molecular Genetics, MD Anderson Cancer
                  Center, 1515 Holcombe-Box 45, Houston, TX 77030-4095, USA
FEATURES          source
                  1..1451
                  /organism="Homo sapiens"
```

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xq22.1-q22.3"
/tissue_type="pooled germ cell tumors"
/note="genomic sequence deposited in GenBank Accession
Number AL049631; derived from dbEST Accession Number
A1964078"
1..1451
/gene="ESX1L"
44..1264
/gene="ESX1L"
/note="ESX1L; expressed in placenta and testis"
/codon_start=1
/product="ESX1L"
/protein_id="AA062141.1"
/db_xref="GI:21629646"
/translation="MESLRGYTHSDIGYSLAVGSDIEVDNDEKLTVSLMARGEDB
ENTRSKPEYGTAEENNVTGEGSVPSDDDBEGGSEPPQOEPPLTPEQOEPP
LLELRKOEERPPQTVEGPQPAEGPOTAEQPPERKRRRTAFQFOLLENPDE
SOYPRVVARERHAAHLNLTEDRVQVFRKAKKGNQVYLMRLTATDLAHPDMF
LGATYAAPALDPALCVHLVPLRPVLPVPPMRPRPMVMPRPDP1AHPMPLAV
PPGSRMAVPDPGRMAVPVPPMPMAVPVPPMPMAVPVPPMPMAVPVPPMPMAVP
PMARVPPGPMAVLPVPPGPMAVLPVPPGPMAVLPVPPGPMAVLPVPPGPMAVLP
MAPVINSYACPPF"
458..640
/gene="ESX1L"
/note="Region: homeodomain"
1432..1437
/gene="ESX1L"
ORIGIN
Query Match      17.9%; Score 103; DB 9; Length 1451;
Best Local Similarity 59.3%; Pred. No. 6.9e-12;
Matches 194; Conservative 0; Mismatches 130; Indels 3; Gaps 1;
QY 246 AGGAGCTCGGAGAGAGCCGCGCGGAGAGAGCGGCCAGGCGGCATGAGAG 305
DB 378 AGCCGCCCAAGACGACCGTGAAGAGGGCCACAGCCGCGGAGGAGCCCAAAACCGTGAAG 437
QY 306 GTCCGAGGCCGAGAACATGACCAAGAACTCGGCGCAGAACTTACGCTGTTGACAG 365
DB 438 GACACAGGCCCCAGAGAGAA---ACGCGCGCGCGCACCGCTTACGACAGTTTCAGC 494
QY 366 TGAAGAGCTGGAAGATGTTTCCGACACATCAATACCTGATGTGCCACAAAGAGG 425
DB 495 TCAGAGAGCTAGAGAACTTTTCGATGAATCTCAATATCCGACGTTGTGGCGGAGAA 554
QY 426 AACTGCCGAAACTAGGTGTGACTGAAGACAAAGTGGGGTGTGTTAAGATPAAA 485
DB 555 GACTTGCAGACAGCCCTGAATTTGACTGAAGACAGAGTGAAGGTTTGTTCAAGACAGAA 614
QY 486 GGGCCAGATGTAAGGCGACATCAGAGAGAAATTAATGCTGCCAATGAATACGCTGACC 545
DB 615 GAGCCAAAGTGAAGAAAGAAATCAGAGGCTCTATGTTGAAGAAACACTGCTACTGAGACC 674
QY 546 CAGAGCAGCTGTGCTACATCGTCGTGG 572
DB 675 TGGCCCAACCTTGGACATGTTCTTGG 701
RESULT 15
AX834967          1518 bp      DNA      linear      PAT 15-DEC-2003
LOCUS             Homo sapiens
DEFINITION        Sequence 2091 from Patent EP1347046.
ACCESSION         AX834967
VERSION           AX834967.1 GI:39921102
KEYWORDS
SOURCE            Homo sapiens (human)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE         1
```


Alignment Scores:	1.38e-73	Length:	19
Pred. No.:		Matches:	19
Score:	1023.00	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%		

Query Match: 97.43% Indels: 0
 DB: 9 Gaps: 0

US-09-867-753-1 (1-577) x US-09-764-864-1155 (1-194)

QY 2 CCAACATCAGGCGCTCCAGCATGGCGCTGCTGCTCCAGACACCGTGTCTACTGC 61
 DB |||||
 DB 4 ProthSergLYAlaProAlaMetAlaArgSerLeuValHisAspThrValPheTyrCys 23
 QY 62 CTGAGTGTATACAGGTTAAATAATAGCCCACTCAGCTGGGGGAGCATCAACGCA 121
 DB |||||
 DB 24 LeuSerValTyrGlnValIlysIleSerProThrProGlnLeuGlyAlaIaIaSerAla 43
 QY 122 GAAGGCATGTTGGCCAGAGAGCTCCAGGCTCATGGTAATATGAACCTGAGGGCGGT 181
 DB |||||
 DB 44 GlnGlyHisValGlyGlnGlyAlaProGlyLeuMetGlyAsnMetAsnProGlnGlyGly 63
 QY 182 GTGAACCAAGAAAGCGCATGAACCGCATGGCGCATGATCCCGAGGGCGCGGTGA 241
 DB |||||
 DB 64 ValAsnHisGlyAsnGlyMetAsnArgAspGlyGlyMetIleProGlnGlyGlyGly 83
 QY 242 AACCAAGAGCTCGGAGAGAGCGGAGCGCCCGCGGAGAGAGCGGCGGAGCGCATG 301
 DB |||||
 DB 84 AsnGlnGlnProArgGlnGlnProGlnProProProGlnGlnProAlaGlnAlaIaMet 103
 QY 302 GAGGTCGCGAGCGCCAGAAACATGACGACGCAACTCGGCGAGCAAGTTACGCTGTTG 361
 DB |||||
 DB 104 GlnGlyProGlnProGlnProGlnMetGlnProArgThrArgThrIlySphThrLeuLeu 123
 QY 362 CAGGTGGAGAGCTGGAAGAGTGTTCCTCCAGACACTCAATACCTGATGTGCCCAACA 421
 DB |||||
 DB 124 GlnValGlnGlnLeuGlnSerValPheArgHisThrGlnIlyProAspValProThrArg 143
 QY 422 AGGGAACCTGCGCAAAACTTAGTGTGACTGAGAGCAAGAAAGTGGGGTTGTTAAGAT 481
 DB |||||
 DB 144 ArgGlnLeuAlaGlnAsnLeuGlyValThrGlnAspIysValArgValIlePheIysAsn 163
 QY 482 AAAAGGCCAGATGTAGGCGCATGACAGAGAAATTAATGCTGCCAATGAATACGTGCT 541
 DB |||||
 DB 164 LysArgAlaArgCysArgArgHisGlnArgGlnLeuMetLeuAlaAsnGlnLeuArgAla 183
 QY 542 GACCCAGAGACTGTGTCTTACATCGCTGTGGAC 574
 DB |||||
 DB 184 AspProAspAspCysValTyrIleValValAsp 194

RESULT 2
 US-09-764-864-11568
 ; Sequence 1568, Application US/09764864
 ; Patent No. US20020132753A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT223
 ; CURRENT APPLICATION NUMBER: US/09/764,864
 ; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 1792
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1568
 ; LENGTH: 194
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (16)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (21)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (22)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE

LOCATION: (25)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (157)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; US-09-764-864-11568

Alignment Scores:
 Pred. No.: 2.9e-71 Length: 194
 Score: 994.00 Matches: 186
 Percent Similarity: 97.38% Conservative: 0
 Best Local Similarity: 97.38% Mismatches: 5
 Query Match: 94.67% Indels: 0
 DB: 9 Gaps: 0

US-09-867-753-1 (1-577) x US-09-764-864-11568 (1-194)

QY 2 CCAACATCAGGCGCTCCAGCATGGCGCTGCTGCTCCAGACACCGTGTCTACTGC 61
 DB |||||
 DB 4 ProthSergLYAlaProAlaMetAlaArgSerLeu**HisAspThrVal*****Cys 23
 QY 62 CTGAGTGTATACAGGTTAAATAATAGCCCACTCAGCTGAGGGGAGCATCAACGCA 121
 DB |||||
 DB 24 Leu***ValTyrGlnValIlysIleSerProThrProGlnLeuGlyAlaIaIaSerAla 43
 QY 122 GAAGGCATGTTGGCCAGAGAGCTCCAGGCTCATGGTAATATGAACCTGAGGGCGGT 181
 DB |||||
 DB 44 GlnGlyHisValGlyGlnGlyAlaProGlyLeuMetGlyAsnMetAsnProGlnGlyGly 63
 QY 182 GTGAACCAAGAAAGCGCATGAACCGCATGGCGCATGATCCCGAGGGCGCGGTGA 241
 DB |||||
 DB 64 ValAsnHisGlyAsnGlyMetAsnArgAspGlyGlyMetIleProGlnGlyGlyGly 83
 QY 242 AACCAAGAGCTCGGAGAGAGCGGAGCGCCCGCGGAGAGAGCGGCGGAGCGCATG 301
 DB |||||
 DB 84 AsnGlnGlnProArgGlnGlnProGlnProProProGlnGlnProAlaGlnAlaIaMet 103
 QY 302 GAGGTCGCGAGCGCCAGAAACATGACGACGCAACTCGGCGAGCAAGTTACGCTGTTG 361
 DB |||||
 DB 104 GlnGlyProGlnProGlnProGlnMetGlnProArgThrArgThrIlySphThrLeuLeu 123
 QY 362 CAGGTGGAGAGCTGGAAGAGTGTTCCTCCAGACACTCAATACCTGATGTGCCCAACA 421
 DB |||||
 DB 124 GlnValGlnGlnLeuGlnSerValPheArgHisThrGlnIlyProAspValProThrArg 143
 QY 422 AGGGAACCTGCGCAAAACTTAGTGTGACTGAGAGCAAGAAAGTGGGGTTGTTAAGAT 481
 DB |||||
 DB 144 ArgGlnLeuAlaGlnAsnLeuGlyValThrGlnAspIys***ArgValIlePheIysAsn 163
 QY 482 AAAAGGCCAGATGTAGGCGCATGACAGAGAAATTAATGCTGCCAATGAATACGTGCT 541
 DB |||||
 DB 164 LysArgAlaArgCysArgArgHisGlnArgGlnLeuMetLeuAlaAsnGlnLeuArgAla 183
 QY 542 GACCCAGAGACTGTGTCTTACATCGCTGTGGAC 574
 DB |||||
 DB 184 AspProAspAspCysValTyrIleValValAsp 194

RESULT 3
 US-09-867-753-2
 ; Sequence 2, Application US/09867753
 ; Patent No. US20020046764A1
 ; GENERAL INFORMATION:
 ; APPLICANT: WEISS, BERTRAM
 ; APPLICANT: GESSERICK, CHRISTOPH
 ; APPLICANT: HAENDLER, BERNARD
 ; TITLE OF INVENTION: HUMAN PEM AS A TARGET FOR BIRTH CONTROL AND TREATMENT
 ; FILE REFERENCE: SCH-1810
 ; CURRENT APPLICATION NUMBER: US/09/867,753
 ; CURRENT FILING DATE: 2001-05-31
 ; PRIOR APPLICATION NUMBER: DE 10027170.7
 ; PRIOR FILING DATE: 2000-05-31
 ; NUMBER OF SEQ ID NOS: 6

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 184
TYPE: PRF
ORGANISM: Homo sapiens
US-09-867-753-2

Alignment Scores:

Pred. No.:	1,26e-70	Length:	184
Score:	986.00	Matches:	184
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	93.90%	Indels:	0
DB:	9	Gaps:	0

US-09-867-753-1 (1-577) x US-09-867-753-2 (1-184)

```
QY 23 ATGGCGCGTTCCTGCTCCACGACCGGTGTTCTACTGCTGATATACAGGTTAA 82
DB 1 MetAlaArgSerLeuValHisAspThrValPheTYrCysLeuSerValTYrGlnValLys 20
QY 83 ATAAAGCCCAACCTCAGCTGGGGGAGCATCAAGGCGGAAGGCCATTTGGCCAGA 142
DB 21 TLeSerProThrProGlnLeuGlyAlaAlaSerSerAlaGlnGlyAlaValGlnGlnGly 40
QY 143 GCTCCAGGCGCTCATGGGTATATGAACCTTGAGGGCGGTGTGAACCAAGAGACGGCATG 202
DB 41 AlAPrProGlyLeuMetClyAsnMetAsnProGlnGlnGlyValAlaAsnHisGlnLysnGlyMet 60
QY 203 AACCGGATGGCGGCGATGATCCCGAGGGCGGCGGTGAACCAAGAGACCGCTCGGACGAG 262
DB 61 AsnAlaGspArgGlyGlyMetIleProGlnGlnGlyGlyGlyAlaAsnGlnGlnProArgGlnGln 80
QY 263 CCGAGCGCCCGCGGAGGAGCGCGGCGGCGCATGAGAGGTCCGAGCCGAGAAC 322
DB 81 ProGlnProProProGlnProGlnProAlaGlnAlaAlaMetClnGlnGlyProGlnProGlnAsn 100
QY 333 ATGACGACCAAGAACTCGGCGCAAGATTCAAGCTTTGACAGGTGAGAGAGCTGGAAGT 382
DB 101 MetGlnProArgThrArgThrArgThrArgThrArgThrArgThrArgThrArgThrArgThr 120
QY 383 GTTTTCCGACACATCAATATACCTGATGTGTGCCCAAGAGGAACTTGCCGAACTTA 442
DB 121 ValIlePheArgHisThrGlnTYrProAspValProThrArgArgGlnLeuAlaGlnLeu 140
QY 443 GGTGTGACTGAAGACAAAGTGGGGGTTTGGTTTAAGATTAAGAGGCCAGATGTAGCGCA 502
DB 141 GlyValThrGlnAspLysValAlaArgValTyrPheLysAsnLysArgAlaArgCysArgArg 160
QY 503 CATCAGAGAGATTAATGCTCGCAATGAACGCTGACCGCAGACGACTGTGTCTAC 562
DB 161 HisGlnArgGlnLeuMetLeuAlaAsnGlnLeuArgAlaAspProAspAspCysValTYr 180
QY 563 ATCGTGTGTGAC 574
DB 181 IleValValAsp 184
```

RESULT 4

US-10-108-260A-4534
Sequence 4534, Application US/10108260A
Publication No. US20040005560A1
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE
TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
FILE REFERENCE: H1-A0106
CURRENT APPLICATION NUMBER: US/10/108,260A
NUMBER OF SEQ ID NOS: 5458
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4534
LENGTH: 415
TYPE: PRF
ORGANISM: Homo sapiens

US-10-108-260A-4534

Alignment Scores:			
Pred. No.:	1,46e-15	Length:	415
Score:	298.50	Matches:	71
Percent Similarity:	49.21%	Conservative:	22
Best Local Similarity:	37.57%	Mismatches:	39
Query Match:	28.43%	Indels:	57
DB:	15	Gaps:	6

US-09-867-753-1 (1-577) x US-10-108-260A-4534 (1-415)

```
QY 170 CCTGAGGCGGTGTGAACCAAGAGACCGGATGAGCGGATGATCCCGCTCC 226
DB 51 ProGlnTYrGlyTYrGlnAlaGlnLysnValGlnTYrGlnGlnGlySer---ValProSer 69
QY 227 -----GAGGCGCGCGGTGAACCAAGAGACCTCGGACGACCGCAGCCC 271
DB 70 AspAspGlnAspArgGlnGlyGlyGlyGlnHisGlnProGlnGlnGlnGlnGlnPro 89
QY 272 CCG ----- 274
DB 90 ProLeuThrLysProGlnGlnGlnGlnGlnGlnGlnProLeuLeuGlnLysGlnGln 109
QY 275 CCGAGAGAGCGCGCCCGGCGGCGCATGAGAGGTCCGACGCC----- 316
DB 110 GlnGlnLysProProGlnThrThrValGlnGlnGlnProGlnProAlaGlnGlnGlnProGlnThr 129
QY 317 ---GAGAACATGCACCA-----CGAACTCGGCGGCGAGAACTCAAGCTGTG 361
DB 130 AlAGlnGlnGlyProGlnProProGlnArgLysArgArgArgArgThrAlaPheThrGlnPhe 149
QY 362 GAGGAGGAGAGCTGGAAGAGTGTTCGACACATCAATCCGATGTGCCCAAGA 421
DB 150 GlnLeuGlnGlnLeuGlnLysnPhePheAspGlnSerGlnTYrProAspValAlaAlaArg 169
QY 422 AGGGAATGTGCCGAATACTTAAAGTGTGACTGAAGACAAAGTGGCGGTGTTAAGAT 481
DB 170 GlnArgLeuAlaAlaArgLeuAsnLeuThrGlnAspArgValGlnValTyrPheGlnAsn 189
QY 482 AAAAGGCCAAGTGAAGCGGACATCAGAGAGATTAATGCTGCCAATGAACGACTGTGT 541
DB 190 ArgArgAlaLysTyrLysArgAsnGlnArgValLeuMetLeuArgAsnThrAlaThrAla 209
QY 541 ----- 541
DB 210 AspLeuAlaHisProLeuAspMetPheLeuGlnGlyAlaTYrTYrAlaAlaProAlaLeu 229
QY 542 GACCCAGACGACTGTGTCTACATCGTC 568
DB 230 AspProAlaLeuCysValHisLeuVal 238
```

RESULT 5

US-10-295-027-676
Sequence 676, Application US/10295027
Publication No. US200302350A1
GENERAL INFORMATION:
APPLICANT: Afar, Daniel
APPLICANT: Aziz, Natasha
APPLICANT: Ginsberg, Wendy M.
APPLICANT: Gish, Kurt C.
APPLICANT: Glynn, Richard
APPLICANT: Hevez, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan R.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
FILE REFERENCE: 018501-012500US
CURRENT APPLICATION NUMBER: US/10/295,027
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/663,733

QY 224 CCGAGGCGCGGTGGAAC----- 244
DB 215 ProAlaAlaGlyGlyThrGlyThrGluAspArgGluGluLeuLeuGluAspGlu 234
QY 244 ----- 244
DB 235 GluAspGluAspGluGluGluLeuLeuGluAspArgGluGluLeuLeuGluAsp 254
QY 245 -----CAGGAGCTCGGCGAGCAGCCG----- 265
DB 255 AspAlaArgAlaLeuLeuGluGluProAlaArgCysProValAlaAlaThrGlyAlaVal 274
QY 266 -----CAGCCCGCGCGGAGAG 283
DB 275 AlaAlaAlaAlaAlaAlaAlaThrGluGlyGlyLeuSerProGlyGluGlu 294
QY 284 -----CCGCGCCAGGCGCGCATGAGGGTCCGACGCCGAGAAC----- 322
DB 295 LeuLeuLeuHisProGluAspAla-----GluGlyLysAspGlyGluAspSerValCys 312
QY 323 -----ATGACGCCAGCACTCGCGCG 343
DB 313 LeuSerAlaGlySerAspSerGluGluGlyLeuLeuLysArgLysGlnArgGlyArg 332
QY 344 AGCAAGTTCAAGCTGTTCAGGTGAGAGAGCTGAAAGTGTTCGACACACTCAATAC 403
DB 333 ThrThrPheThrSerGlyGlnLeuGluGluArgAlaPheGlnLysThrHisGly 352
QY 404 CCTGATGTCCCAAGAAAGGAACTTCCGAAACTTGGGTGACTGAGACAAAGTG 463
DB 353 ProAspValPheThrArgGluGluLeuAlaMetArgLeuAspLeuThrGluAlaArgVal 372
QY 464 CGGGTTGGTTAAGATAAAGCGCCAGATGATGACCATCAGAGA 511
DB 373 GlnValTTPPheGlnMetArgAlaLysTTPArgLysArgGluLys 388
RESULT 7
US-10-161-051-88
; Sequence 88, Application US/10161051
; Publication No. US20030152945A1
; GENERAL INFORMATION:
; APPLICANT: Peter Deak
; APPLICANT: David M Glover
; APPLICANT: Carol Mdgley
; TITLE OF INVENTION: Cell cycle progression proteins
; FILE REFERENCE: CCI-021CP
; CURRENT APPLICATION NUMBER: US/10/161,051
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: GB 0007268.6
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 88
; LENGTH: 1506
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-161-051-88
Alignment Scores:
Pred. No.: 7,88e-07 Length: 1506
Score: 190.00 Matches: 55
Percent Similarity: 43.86% Conservative: 20
Best Local Similarity: 33.16% Mismatches: 60
Query Match: 18.10% Indels: 36
DB: 14 Gaps: 7
US-09-867-753-1 (1-577) x US-10-161-051-88 (1-1506)
QY 92 AACCTCAGCTGGGGGACATCAACGCAAGAGCATGTTGGCAAGAGCTCCAGGC 151
DB 1123 ThrProProGlyGlyGlyProAlaGlyAlaGlyAlaLeuGlnProGlyGlySerGly 1142
QY 152 CTCATG-----GTAATATG-----AACCT-----GAGGCGGTGTG 184

DB 1143 SerSerGlySerAspGlyAsnMetSerSerAsnProAsnSerSerAsnThr 1162
QY 185 AACCAAGAAAGCGCATGACCGCATGGCGC----- 217
DB 1163 ThrHisSerAsnGlyHisAsnThrAsnSerGlySerGlyCysGlyAspSerAlaGly 1182
QY 218 -----ATGATCCCGAGGCGGCGGTGAAACAGAGAGCT 253
DB 1183 SerGlyArgLeuSerLeuProAlaLeuSerProAspSerGlySer----- 1197
QY 254 CCGCAGCAGCCGAGCGCCCGCGGAGAGAGAGCCGCGCATGAGAGGTCCGAG 313
DB 1198 ArgAspSerArgSerProAspAlaAspAlaAsnArgMetIleAsp11eGluGlyAsp 1217
QY 314 CCGAGAACATG-----CAGCCACGAACCTCG-----CGACAGAGTTCAAGCTG 358
DB 1218 SerGluSerGlnAspSerAspGlnProLysPheArgArgAsnArgThrPheSerPro 1237
QY 359 TTGACGTGAGAGAGCTGGAAGTGTTCGACACACTCAATACCTGATGCGCACA 418
DB 1238 GluGlnLeuAspGluLeuGluGlyGluPheAspLysSerHisGlyProCysValAsnThr 1257
QY 419 AGAAGGAACTTCCGAAACTTAACTGACTGAAGACAAAGTGGGTTGTTAAG 478
DB 1258 ArgGluLysLeuAlaAlaArgThrAlaLeuSerGluAlaArgValGlnValTTPPheSer 1277
QY 479 AATAAAGGCCAGATGTGCGGACATCAGAGA 511
DB 1278 AsnArgArgAlaLysTTPArgArgHisGlnArg 1288
RESULT 8
US-10-221-625-88
; Sequence 88, Application US/10221625
; Publication No. US20040033942A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
; APPLICANT: LU, Dying Aina M.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: AZIMZAI, Valda
; APPLICANT: BANDMAN, Olga
; APPLICANT: TANG, Y. Tom
; APPLICANT: MATHUR, Preete
; APPLICANT: SHAH, Purvi
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: REDDY, Roopa
; TITLE OF INVENTION: TRANSCRIPTION FACTORS
; FILE REFERENCE: PF-0761 PCT
; CURRENT APPLICATION NUMBER: US/10/221,625
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PERL Program
; SEQ ID NO 88
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040033942A1 3717139CD1
US-10-221-625-88
Alignment Scores:
Pred. No.: 1,07e-06 Length: 215
Score: 187.50 Matches: 37
Percent Similarity: 58.89% Conservative: 16
Best Local Similarity: 41.11% Mismatches: 30
Query Match: 17.86% Indels: 7
DB: 12 Gaps: 1

```

US-09-867-753-1 (1-577) x US-10-221-625-88 (1-215)
QY 269 CCCCCCGAGAGAGCCGCGCC-----CAGCGCCCATGAGAGGT 307
Db 9 PProPProValSerProGluLeuLysAspArgLysGluAspAlaLysGlyMetGluAsp 28
QY 308 CCGCAGCCCGAGAGATGACCCAGCAACTCGGCGCAGCAAGATTCAAGCTGTTGACAGGTG 367
Db 29 GluLysGlnThrLysLeuLysGlnArgSerArgThrAspMetThrLeuGlnGluLeu 48
QY 368 GAGAGACTGGAAGAGTGTTCCTCCACACTCACTACCTGATGTCGCCCAAGAGAGAA 427
Db 49 AsnGlnLeuGlnArgLeuPheAspGlnThrHisLysProAspAlaPheMetArgGlnGlu 68
QY 428 CTTCGCGAAACTTATGATGCTGAGCAAGCAAGTGGGTGTTTAAAGATAAAG 487
Db 69 LeuSerGlnArgLeuLysLeuSerGlnAlaArgValGlnValTrpPheGlnAsnArgArg 88
QY 488 GCCAGATGTAGCGCAGCATCAGAGAAATTA 517
Db 89 AlaLysCysArgLysGlnGlnGlnGlnLeu 98

RESULT 9
US-10-301-822-165
; Sequence 165, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: BURGART, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MEMO1-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165
; LENGTH: 217
; TYPE: PRN
; ORGANISM: Homo Sapiens
US-10-301-822-165

Alignment Scores:
Pred. No.: 1,86e-06 Length: 217
Score: 184.50 Matches: 49
Percent Similarity: 44.97% Conservative: 27
Best Local Similarity: 28.99% Mismatches: 54
Query Match: 17.57% Indels: 39
DB: 14 Gaps: 5

US-09-867-753-1 (1-577) x US-10-301-822-165 (1-217)
QY 113 TCAGAGCGCAGAGAGCCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTAAATGAACCTT 172
Db 2 ThrSerSerTyrGlnHisValLeuGlnArgGlnProAlaLeuGlnLysGlnArgLeuAspSer 21
QY 173 GAGGCG-----GGTGTGAACCGCAGGAACGCG 139
Db 22 ProGlyAsnLeuAspThrLeuGlnAlaLysLysAsnPheSerValSerHisLeuLeuAsp 41

```

```

QY 200 ATGAACCGCATGCGCGCATGATCC-----GAGGCGGC 235
Db 42 LeuGlnGlnAlaGlyAspMetValAlaAlaGlnAlaAspGluAsnValGlnGlnAlaGly 61
QY 236 -----GGTGAACACAGAGAGCTCGGCGAGCGCG 265
Db 62 ArgSerLeuLeuGlnSerProGlyLeuThrSerGlySerAspThrProGlnGlnAsp--- 80
QY 266 CAGCCCCCGAGAGAGAGCGGCCCGAGCGGCATGAGAGGGTCCGACGCCCGAGAACATG 325
Db 81 -----AsnAspGlnLeuAsnSerGlnGlnLysLysLysArg 92
QY 326 CAGCCACGAATCTCGCGCGCAGAACTTACAGCTTTCACAGTGGAGAGAGCTGGAAGTGT 385
Db 93 LysGlnArgArgAsnArgThrThrPheAsnSerSerGlnLeuGlnAlaLeuGlnArgVal 112
QY 386 TTCCGCACACATCAATACCTGATGTCGCCCAAGAGGAAGCTTCCGCAAACTTAGGT 445
Db 113 PheGlnArgThrHisLysProAspAlaPheValArgGlnAspLeuAlaArgArgValAsn 132
QY 446 GTGACTGAAGACAAAGTGGCTTTGTTTAAAGATAAAGGCGCAGATGTAGCGACAT 505
Db 133 LeuThrGlnAlaArgValGlnValTrpPheGlnAsnArgArgAlaLysPheArgArgAsn 152
QY 506 CAGAGAGAAATTAATGCTCGCCCAATGA 532
Db 153 GluArg---AlaMetLeuAlaAsnLys 160

RESULT 10
US-10-301-822-167
; Sequence 167, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: BURGART, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MEMO1-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 245
; TYPE: PRN
; ORGANISM: Homo Sapiens
US-10-301-822-167

Alignment Scores:
Pred. No.: 1,88e-06 Length: 245
Score: 184.50 Matches: 49
Percent Similarity: 44.97% Conservative: 27
Best Local Similarity: 28.99% Mismatches: 54
Query Match: 17.57% Indels: 39
DB: 14 Gaps: 5

US-09-867-753-1 (1-577) x US-10-301-822-167 (1-245)
QY 113 TCAGAGCGCAGAGAGCCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTAAATGAACCTT 172
Db 113 TCAGAGCGCAGAGAGCCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTAAATGAACCTT 172

```



```

Db      121 ArgThrAsnPherThleuGlInleuAenGluLeuGluArgLeuPheAspGluThrHis 140
QY      401 TACCCTGATGTCGCCACAGAGGAAGGAACTTGCCGAAATCTAGTGTGACTGAGACAAA 460
        |||||
Db      141 TyrProAspAlaPheMetArgGluGluLeuSerGlnArgLeuGluLeuSerGlnAlaArg 160
QY      461 GTCCGGGTTGGTTTAAAGATTAAGGGCCAGATGTAGCCGACATAGAGAGA----- 514
        |||||
Db      161 ValGlnValIleTyrPheGlnAsnArgAlaIleCysArgGluGlnGluAsnMetHis 180
QY      515 -----TTAATGCTCGCCCATGAACTACGTGCT 541
        |||||
Db      181 LysGlyValIleLeuGlyThrAlaAsnHisLeuAspAla 193

RESULT 13
US-10-158-160A-11
; Sequence 11, Application US/10158160A
; Publication No. US20030059805A1
; GENERAL INFORMATION:
; APPLICANT: RAPFOLD-HOERBRAND, GUDRUN
; APPLICANT: RAO, ERCOLE
; TITLE OF INVENTION: HUMAN GROWTH GENE AND SHORT STATURE GENE REGION
; FILE REFERENCE: 108351-00004
; CURRENT APPLICATION NUMBER: US/10158,160A
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 09/147,699
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: PCT/EP97/05355
; PRIOR FILING DATE: 1997-09-29
; PRIOR APPLICATION NUMBER: 60/027,633
; PRIOR FILING DATE: 1996-10-01
; PRIOR APPLICATION NUMBER: EP/97100583.0
; PRIOR FILING DATE: 1997-01-16
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-158-160A-11

Alignment Scores:
Pred. No.: 8,31e-06 Length: 292
Score: 176.50 Matches: 49
Percent Similarity: 49.71% Conservative: 37
Best Local Similarity: 28.32% Mismatches: 70
Query Match: 16.81% Indels: 17
DB: 14 Gaps: 4

US-09-867-753-1 (1-577) x US-10-158-160A-11 (1-292)
QY      44 GACACGGTGTTCACGCGCTGATATACACGATTAATAAGCCCACTACCTGACTG 103
        |||||
Db      31 AspSerIleThrTyr---ArgGluValLeuGluSerGlyLeuAlaArgSerArgGluLeu 49
QY      104 GGGGACGACATCAAGCCAGAGAGCCATGTGTGCCAAGAGAGCTCCAGGCTCATGGGTAAAT 163
        |||||
Db      50 GlyThrSerAspSerIleuGlnAspIleThrGluGlyGlyGlyHisCysProValHis 69
QY      164 ATGAACCTCAGAGGGCGGTGTGAACACAGAGAACGCGCATGAACCCGATGCC---GGCATG 220
        |||||
Db      70 LeuPheIleAspHisValAspAsnAspGlyGlyLeuLeuGlySerGluPheGlyThrAlaArg 89
QY      221 ATCCCGAGAGGGCGCGGTGGAACACAGAGCTCCGAGACAGCCGACCCCGCGGAG 280
        |||||
Db      90 ValAlaGluIleGlyIleTyrGluCysGlyGlyLeuArgGluValIleGlySerGluAspGlu 109
QY      281 GAGCCGCGCCAGCGCGGCGCATGAGGGTCCGACCCGAGAACATGACGACCACTCGG 340
        |||||
Db      110 Asp-----GlyGlnThrIleuLeuSerGlnArgGlySer 120
QY      341 CGCAGAGAGTTCAAGCTGTTGACAGTGGAGAGCTGGAAGAGTTTCCGACACTCAA 400
        |||||

```

```

Db      121 ArgThrAsnPherThleuGlInleuAenGluLeuGluArgLeuPheAspGluThrHis 140
QY      401 TACCCTGATGTCGCCACAGAGGAAGGAACTTGCCGAAATCTAGTGTGACTGAGACAAA 460
        |||||
Db      141 TyrProAspAlaPheMetArgGluGluLeuSerGlnArgLeuGluLeuSerGlnAlaArg 160
QY      461 GTCCGGGTTGGTTTAAAGATTAAGGGCCAGATGTAGCCGACATAGAGAGA----- 514
        |||||
Db      161 ValGlnValIleTyrPheGlnAsnArgAlaIleCysArgGluGlnGluAsnMetHis 180
QY      515 -----TTAATGCTCGCCCATGAACTACGTGCT 541
        |||||
Db      181 LysGlyValIleLeuGlyThrAlaAsnHisLeuAspAla 193

RESULT 14
US-10-087-192-1983
; Sequence 1983, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1983
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(489)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-10-087-192-1983

Alignment Scores:
Pred. No.: 8,67e-06 Length: 489
Score: 176.50 Matches: 47
Percent Similarity: 49.22% Conservative: 16
Best Local Similarity: 36.72% Mismatches: 44
Query Match: 16.81% Indels: 21
DB: 12 Gaps: 4

US-09-867-753-1 (1-577) x US-10-087-192-1983 (1-489)
QY      173 GAGGGCGGTGTGAACACAGAGGAAGGCGATGAACCGC-----GATGGCGGC 217
        |||||
Db      169 GluGlyAspIleGlyGluGluAspGlyGlyValIleHisSerIleAspGlyIle 188
QY      218 ATGATCCCGAGAGGGCGCGGTGGAACACAGAGCTCCGACAGAGCCGACCCCGCGG 277
        |||||
Db      189 Leu-----GlyAspIleGlyAsnArg-----Leu 196
QY      278 GAGAGCCGCGCCAGCGCGGCGCATGAGGGTCCGACCCGAGAACATGACACCGAAGACT 337
        |||||
Db      197 AspGluGlySerAspValGluSerGluProAspLeuProLeuIleAspArgIleArg 216
QY      338 CGGCGAGCAAGATTCACGCTGTTCAGGTGAGAGAGAGTGAAGTGTTCGACACT 397
        |||||
Db      217 SerArgThrThrPheThrAlaGluGluLeuGluGluLeuGluAlaPheGluArgThr 236
QY      398 CAATACCTGATGTGCCACAGAGGAAGGAACTTGCCGAAATCTAGTGTGACTGAGAC 457
        |||||
Db      237 HisTyrProAspIleIleThrArgGluGluLeuAlaGlnArgHisThrIleThrGluAla 256
QY      458 AAGTGGGGTGTGTTTAAAGATTAAGGGCCAGATGTAGCGGACATCAGAGAGATTA 517
        |||||

```

Db 257 ArgValGlnValTPrpSerAsnArgAlaArgTPrArgLysGlnAlaGly----- 274

Qy 518 ATGCTCGCCATGAACTACGTGCT 541
 Db 275 -----AlaAsnGlnLeuAlaAla 280

RESULT 15

US-10-245-171A-7
 ; Sequence 7, Application US/10245171A
 ; Publication No. US20030124102A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rudnicki, Michael A.
 ; APPLICANT: Seale, Patrick
 ; TITLE OR INVENTION: Pax-Encoding Vector and Use Thereof
 ; FILE REFERENCE: 50120/006002
 ; CURRENT APPLICATION NUMBER: US/10/245,171A
 ; CURRENT FILING DATE: 2003-03-11
 ; PRIOR APPLICATION NUMBER: US 60/322,923
 ; PRIOR FILING DATE: 2001-09-17
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 479
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-245-171A-7

Alignment Scores:

Pred. No.:	9,496-06	Length:	479
Score:	176.00	Matches:	39
Percent Similarity:	56.04%	Conservative:	12
Best Local Similarity:	42.86%	Mismatches:	38
Query Match:	16.76%	Indels:	2
DB:	14	Gaps:	1

US-09-867-753-1 (1-577) x US-10-245-171A-7 (1-479)

Qy 269 CCCCCCGGAGAGACCCGCGGCGGCGCATGAGGTCGCCAGCCCGAAGACATGCAG 328
 Db 199 ProGlnSerAspGlnGlySerAspIleAspSerGlnProAspLeuProLeuLysArgLys 218
 Qy 329 CCAGCACTCGGCGGAGCAAGTTCAAGCTGTGCGAGGAGGCTGGAAGTGTTC 388
 Db 219 GlnArgArgSerArgThrPheThrAlaGlnGlnLeuGlnLeuGlnAlaPhe 238
 Qy 389 CGACACACTCAATACCTGATGTGCCACAGAGGAACTGCCGAAAATTAGGTGTG 448
 Db 239 GlnArgThrHisTyrProAspIleTyrThrArgGlnGlnLeuAlaGlnAlaLysLeu 258
 Qy 449 ACTGAAGACAAAGTGGGGTTGGTTTAAGATTAAGGGCCAGATGTAGCGACAT--- 505
 Db 259 ThrGlnAlaArgValGlnValTPrpSerAsnArgAlaArgTPrArgLysGlnAla 278
 Qy 506 ---CAGAGAGATTATGCTGCCCAATGAACCTA 535
 Db 279 GlyAlaAsnGlnLeuMetAlaPheAsnHisLeu 289

Search completed: June 16, 2004, 19:05:52
 Job time : 49.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 16, 2004, 18:58:15 ; Search time 19.5 Seconds
(without alignments)
3055.198 Million cell updates/sec

Title: US-09-867-753-1

Perfect score: 1050
Sequence: 1 tccacatcagcgcctcag.....tctacatcgtcgtgactag 577

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 38914 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODE=frame_n2p.model -DEV=1p
-O=/cgn2_1/USPTO_spool_p/US09867753/runat_16062004_163636_27826/app_query.fasta_1.775
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09867753_@CGN_1.1.27_@runat_16062004_163636_27826 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	190.5	18.1	302	4	US-08-957-351-7
2	190.5	18.1	313	4	US-08-957-351-9
3	189.5	18.0	302	4	US-08-957-351-30
4	188	17.9	305	2	US-08-957-351-3
5	183.5	17.5	302	2	US-08-957-351-37
6	181.5	17.3	315	4	US-08-957-351-27
7	176	16.8	247	3	US-09-129-888-2
8	174	16.6	234	4	US-09-105-470B-4
9	173	16.5	185	4	US-09-976-594-443
10	171.5	16.3	289	4	US-09-976-594-945
11	169.5	16.1	271	4	US-08-957-351-26
12	169	16.1	241	4	US-09-636-215-842

13	169	16.1	241	4	US-09-685-166A-842	Sequence 842, App
14	167	15.9	60	4	US-08-754-477A-18	Sequence 18, Appl
15	167	15.9	60	4	US-08-754-477A-19	Sequence 19, Appl
16	164.5	15.7	132	4	US-08-754-477A-135	Sequence 135, App
17	164.5	15.7	271	4	US-08-754-477A-2	Sequence 2, Appl1
18	164.5	15.7	271	4	US-08-754-477A-5	Sequence 5, Appl1
19	163	15.5	60	4	US-08-754-477A-11	Sequence 11, Appl
20	162	15.4	60	4	US-08-754-477A-10	Sequence 10, Appl
21	162	15.4	234	4	US-09-105-470B-2	Sequence 2, Appl1
22	161	15.3	60	4	US-08-754-477A-14	Sequence 14, Appl
23	161	15.3	240	4	US-09-636-735A-2	Sequence 2, Appl1
24	161	15.3	240	4	US-09-636-735A-12	Sequence 12, Appl
25	160	15.2	60	2	US-08-775-009-38	Sequence 38, Appl
26	160	15.2	60	4	US-08-754-477A-13	Sequence 13, Appl
27	160	15.2	60	4	US-08-754-477A-15	Sequence 15, Appl
28	160	15.2	60	4	US-08-754-477A-17	Sequence 17, Appl
29	160	15.2	60	4	US-08-754-477A-20	Sequence 20, Appl
30	160	15.2	99	4	US-09-031-962D-24	Sequence 24, Appl
31	159	15.1	60	4	US-08-754-477A-16	Sequence 16, Appl
32	159	15.1	436	2	US-08-958-642-4	Sequence 4, Appl1
33	159	15.1	436	3	US-08-778-394-2	Sequence 2, Appl1
34	159	15.1	436	3	US-08-778-423A-4	Sequence 4, Appl1
35	158.5	15.1	271	4	US-08-754-477A-121	Sequence 121, App
36	158.5	15.1	271	4	US-08-754-477A-125	Sequence 125, App
37	158	15.0	60	4	US-08-754-477A-7	Sequence 7, Appl1
38	157.5	15.0	271	4	US-08-754-477A-131	Sequence 131, App
39	157	15.0	60	4	US-08-754-477A-12	Sequence 12, Appl
40	157	15.0	60	4	US-08-754-477A-23	Sequence 23, Appl
41	156.5	14.9	302	2	US-08-203-532F-4	Sequence 4, Appl1
42	156.5	14.9	302	3	US-08-950-860-16	Sequence 16, Appl
43	156.5	14.9	302	3	US-09-078-465-4	Sequence 4, Appl1
44	156.5	14.9	302	5	PCT-US95-01882A-4	Sequence 4, Appl1
45	155	14.8	434	2	US-08-710-249-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-957-351-7
Sequence 7, Application US/08957351
Patent No. 6306586
GENERAL INFORMATION:
APPLICANT: Murray, Jeffrey C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
NUMBER OF SEQUENCES: 33
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-351-7

Alignment Scores:

Pred. No.:	Length:
4.23e-10	302
Score: 190.50	Matches: 49
Percent Similarity: 50.70%	Conservative: 23
Best Local Similarity: 34.51%	Mismatches: 45
Query Match: 18.14%	Indels: 25
DB: 4	Gaps: 5

US-09-867-753-1 (1-577) x US-08-957-351-7 (1-302)

QY 98 CAGCTGGGGGAGCATCAAGCGGAGGCGCATGTTGGCCAAAGAGCTCCAGGCTCATG 157
DB 2 GlnPheGlyLeuLeuSerGluAlaGluAlaArg-----SerProAlaLeuSer 17
QY 158 GGTAAATATGAACCCCTGAGGGCGGTGTGAACACAGAGAACGGCATGAACCGCATGGCGGC 217
DB 18 LeuSer-----AspAlaGlyThrProHisPro----- 26

QY 218 ATGATCCCGGAGGCGCGGTGTGAACACAGAGCTTCGGCAGACCGCCGACGCCCGCG 277
DB 27 GlnLeuProGluHisGlyCysLysGlyGlnGlnHisSerAsp----- 40
QY 278 GAGAGCCGGCGGAGGCGCATGAGAGGTCGCGACCGCGAGAAC-----ATG 325
DB 41 SerGlnLysAlaSerLysLeuProGlyGlySerProGluAspLysLeuLys 60

QY 326 CAGCCAGCACTCGGCGGAGGAGTTACGCGTGTGCAAGTGGAGGAGGAGGAGGAGGAGT 385
DB 61 LysGlnArgArgGlnArgThrHisPheThrSerGlnGlnLeuGlnGlnLeuGlnAlaThr 80
QY 386 TTCGACACACTCAATACCTGATGTGCCACAGAGGAACTTCCGCAAACTTAGGT 445
DB 81 PheGlnArgAsnArgLysProAspMetSerThrArgGlnGlnLeuAlaValAlaTrpThrAsn 100

QY 446 GTGACTGAAGACAAAGTGGTTGTTTAAGATAAAGGCGCAGATGAGCGACAT 505
DB 101 LeuThrGlnAlaArgValArgValTrpPheLysAsnArgArgAlaLysTrpArgLysArg 120
QY 506 CAGAGA 511
DB 121 GlnArg 122

RESULT 2
US-08-957-351-9
Sequence 9, Application US/08957351
Patent No. 6306586

GENERAL INFORMATION:
APPLICANT: Semina, Elena
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston

STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 313 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-351-9

Alignment Scores:
Pred. No.: 4.27e-10
Score: 190.50
Percent Similarity: 50.70%
Best Local Similarity: 34.51%
Query Match: 18.14%
DB: 4
Length: 313
Matches: 49
Conservative: 23
Mismatches: 45
Indels: 25
Gaps: 5

US-09-867-753-1 (1-577) x US-08-957-351-9 (1-313)

QY 98 CAGCTGGGGGAGCATCAAGCGGAGGCGCATGTTGGCCAAAGAGCTCCAGGCTCATG 157
DB 2 GlnPheGlyLeuLeuSerGluAlaGluAlaArg-----SerProAlaLeuSer 17
QY 158 GGTAAATATGAACCCCTGAGGGCGGTGTGAACACAGAGAACGGCATGAACCGCATGGCGGC 217
DB 18 LeuSer-----AspAlaGlyThrProHisPro----- 26

QY 218 ATGATCCCGGAGGCGCGGTGTGAACACAGAGCTTCGGCAGACCGCGACGCCCGCGCG 277
DB 27 GlnLeuProGluHisGlyCysLysGlyGlnGlnHisSerAsp----- 40
QY 278 GAGAGCCGGCGGAGGCGCATGAGAGGTCGCGACCGCGAGAAC-----ATG 325
DB 41 SerGlnLysAlaSerLysLeuProGlyGlySerProGluAspLysLeuLys 60

QY 326 CAGCCAGCACTCGGCGGAGGAGTTACGCGTGTGCAAGTGGAGGAGGAGGAGGAGGAGT 385
DB 61 LysGlnArgArgGlnArgThrHisPheThrSerGlnGlnLeuGlnGlnLeuGlnAlaThr 80
QY 386 TTCGACACACTCAATACCTGATGTGCCACAGAGGAACTTCCGCAAACTTAGGT 445
DB 81 PheGlnArgAsnArgLysProAspMetSerThrArgGlnGlnLeuAlaValAlaTrpThrAsn 100

QY 446 GTGACTGAAGACAAAGTGGTTGTTTAAGATAAAGGCGCAGATGAGCGACAT 505
DB 101 LeuThrGlnAlaArgValArgValTrpPheLysAsnArgArgAlaLysTrpArgLysArg 120
QY 506 CAGAGA 511
DB 121 GlnArg 122

RESULT 3
US-08-957-351-30
Sequence 30, Application US/08957351
Patent No. 6306586

GENERAL INFORMATION:
APPLICANT: Semina, Elena
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA


```

ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-351-30

Alignment Scores:
Pred. No.: 5,27e-10 Length: 302
Score: 189.50 Matches: 49
Percent Similarity: 50.00% Conservative: 22
Best Local Similarity: 34.51% Mismatches: 46
Query Match: 18.05% Indels: 25
Gaps: 5
DB: 4

US-09-867-753-1 (1-577) x US-08-957-351-30 (1-302)
QY 98 CAGCTGGGGGAGCATCAAGCCGAGAGGCGCATGTTGGCCAGAGCTCCAGGCGCTCATG 157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 GlnpneGlyLeuLeuSerGlnLagLualAArg-----ThrProAlaLeuSer 17

QY 158 GGTAAATATGAACCTGAGGGCGGCTGTGAACACAGAGAACCGCATGAACCGCATGGCGGC 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18 LeuSer-----AspLagLylThrProHisPro----- 26

QY 218 ATGATCCCGGAGGGCGCGCGGTGGAACACAGAGACCTCGGACAGCGCCGACGCCGCCG 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27 GlnpneProGlnHisGlyCyulysGlyGlnGlnHisSerAsp----- 40

QY 278 GAGGAGCGCGCCGAGGCGCGCATGAGAGGCGCGACGCCCGAGAAC-----ATG 325
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 SerGlnLysAlaSerHisSerLeuProGlyGlySerProGlnAspGlySerLeuLysLys 60

QY 326 CAGCCAGCAACTCGGCGGACAGAAATTCAAGCTGTTCCAGGTGAGAGAGCTGAGAAAGTTT 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 LysGlnAArgArgGlnAArgThrHisPheThrSerGlnGlnLeuGlnGlnLalAThr 80

QY 386 TTCGACACATCTCATATACCTGTATGTGCCACAGAAAGGGAACCTTGCCGAGAACTTAAGT 445
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 PheGlnAArgAsnAArgTyrProAspMetSerThrAArgGlnGlnLalAValItrPThrAsn 100

QY 446 GTGATCTGAAGACAAAGTGCGGGTGTGTTTAAGATPAAAGGGCCAGATGTAGCGCAT 505
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 LeuThrGlnAlaArgValAArgValItrPheLysAsnAArgArgAlaLysItrPArgLysArg 120

QY 506 CAGAGA 511
   ||| |||
Db 121 GlnAArg 122

RESULT 4
US-08-957-351-3
; Sequence 3, Application US/08957351
; Patent No. 6306586
; GENERAL INFORMATION:
; APPLICANT: Semina, Elena

```

APPLICANT: Murray, Jeffrey C.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
 TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
 NUMBER OF SEQUENCES: 33
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FOLEY, HONG & ELIOT LLP
 STREET: One Post Office Square
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109-2170
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/957,351
 FILING DATE: 24-OCT-1997
 CLASSIFICATION: 415
 ATTORNEY/AGENT INFORMATION:
 NAME: Arnold, Beth E.
 REGISTRATION NUMBER: 35,430
 REFERENCE/DOCKET NUMBER: UIA-024.01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-832-1000
 TELEFAX: 617-832-7000
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 302 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 JS-08-957-351-3

Alignment Scores:			
Pred. No.:	7, 32e-10	Length:	302
Score:	188.00	Matches:	41
Percent Similarity:	59.41%	Conservative:	19
Best Local Similarity:	40.59%	Mismatches:	31
Query Match:	17.90%	Indels:	10
DB:	4	Gaps:	2
US-09-867-753-1 (1-577) x US-08-957-351-3 (1-302)			
QY	221 ATCCCGAGGCGCGCGGTGAAAACGAGAGCTCGGCGAGCGACGCCGCGCGGAG	280	
Db	28 LennProGlnHhGILCYelyselYglnIuhIseerhp-----Ser 41		
QY	281 GAGCCCGCGCGCGCGCATGAGGAGGTCCGCGACGCCGAGAAC-----ATGCAG	328	
Db	42 GlnUlyalIseRlIseRleuProIylIseRProGlnaepGlyseRleuYlYlYs 61		
QY	329 CCACGAACTCGGCGCGACGAAAGTTCA CGCTGTTGCAGGTGAGAGAGCTGGAAAGTGTTC	388	
Db	62 GlnArGArGlnArGThrhIspherhserGlnGlnleuGlnIuLeuGlnAlaThrPhe 81		
QY	389 CGACACACTCAATACCTCGATGATGCCCAAGAAAGGAACTTGGCCGAAAACTTAGTGTG	448	
Db	82 GlnArGAsArGlyrProAspMetserThrhArgIuGlnIuLeaIaValTrpThrAsnLeu 101		
QY	449 ACTGAAGACAAAGTCGCGGTTTGTTGATTAAAGATAAAGGCGCAATATGAGCGCATCAG	508	
Db	102 ThrGlnAlaArValArGValTrpPheIysAsnArGAlaIaLyrTrpArGlyArGln 121		
QY	509 AGA 511		
Db	122 Arg 122		

RESULT 5
US-08-775-009-37
; Sequence 37, Application US/0875009
; Patent No. 5935783


```

Db      67 rArgalaglyAlaGlnAsnAspGlnLeuSerThrGlyProArgAlaAlaProGluGlu1 87
Qy      286 GGGCCAGGGGGCCATGAGGGGTCCGCAG-----CCCA 318
Db      87 aglurhLeuAlaGluThrGlnProGluLurghIsleuGlySerTyrLeuLeuAspSerG1 107
Qy      319 GAAC-----ATGCAGCCAGCAAGCTCG-- 340
Db      107 uanenthSerGlyAlaLeuProArgLeuProGlnThrProGlyLeuProGlnLysArgSe 127
Qy      341 -CGCAGAACTTACCGCTGTTCAGAGCTGAGAGCTTGAAAGTGTTCGACACACTCA 399
Db      127 rArgalaaAlaPheSerHisThrGlnValIleGluLeuGluArgLysPheSerHisGlnLys 147
Qy      400 ATACCGCTGATGTGCCCAAGAGGAGGAACTTGCGGAAACTTGAGTGTGATGAGAGCA 459
Db      147 sTyrLeuSerAlaAlaProGluLurghAlaHisleuAlaLysAsnLeuLysLeuThrGluThrG1 167
Qy      460 AGTCGGGGTTTGGTTTAAAGATAAAGGGCCAGATGATGAGCGACATCAGAGAAATTAAAT 519
Db      167 nValylsIleTetPheGlnAsnArgArgTyrLysThrLysArgLysGln----- 183
Qy      520 GCTGCCAATGAACTA 535
Db      184 -LeuSerSerGluLeu 188

```

RESULT 9

```

US-09-976-594-443
; Sequence 443, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 443
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 168861CD1
US-09-976-594-443

```

Alignment Scores:

```

Pred. No.: 1 75e-08 Length: 185
Score: 173.00 Matches: 53
Percent Similarity: 44.83% Conservative: 25
Best Local Similarity: 30.46% Mismatches: 53
Query Match: 16.48% Indels: 43
DB: 4 Gaps: 7

```

```

US-09-867-753-1 (1-577) x US-09-976-594-443 (1-185)
Qy      76 GGTAAATTAAGCCCAACCTCAGCTGGGGGAGCATCAAGCGC----- 120
Db      12 GtlyGlnAsnLysProSerThr-----CysSerPheSerIleGluArgIleuGlyLeuAsp 30
Qy      121 ---AGAGCGCATGTTGGCCA-----AGAGCTCCAGGCTTATGGGTAATTAAGACC 171
Db      31 GlnLysLysAspCysValProLeuMetLysProHisArgPro-TripAlaAspThrCysSe 50
Qy      172 TGAAGCGCGGTGAACACGAGAAAGCGCATGAACCGGATGCGCGCATG-----AT 222
Db      50 rSerSerGly-----LysAspGlyAsnLeuLysLeuHisbVa 62

```

```

Qy      223 CCCCAGAGGGCGCGGTGAAACACGAGAGCTCGGACAGACCGCCCGGAGAGA 282
Db      62 lProAsnProProSerGlyIleSerPheProSerValValaPheIsPromeProGluG1 82
Qy      283 GCGCGCCAGGGGGCCATGAGAGGTCCGACGCCCGAGAAC----- 322
Db      82 uArgAlaSer-----LysTyrGluAsnTyrPheSerAlaSerGluAr 96
Qy      323 -----ATGCAGCCAGCAAGCTCGGCGCAGCAAGTTCAC 354
Db      96 glueSerLeuLysArgGluLeuSerTyrTyrArgGlyArgAlaGArgProArgThrAlaPheTh 116
Qy      355 GCTGTTCAGGTGAGAGAGACTGGAAGTGTTCGACACACTCAATACCTGATGTGCC 414
Db      116 rGlnAsnGlnIleGluValLeuGluAsnValPheArgValaAsnCyTyrProGlyLys 136
Qy      415 CACAGAAAGGAACTTGGCCGAAACTTACGTGATGACTGAGCAAGCAAGTGGTGT 474
Db      136 pIleArgGluAspLeuAlaGlnLysLeuAsnLeuGluAspArgIleGlnIleTetPhe 156
Qy      475 TAAGAAATAAAGGGCCAGATGATGAGCGACATCAGAGAGA 514
Db      156 eGlnAsnArgArgAlaLysLeuLysArgSerHisArgGlu 169

```

RESULT 10

```

US-09-976-594-945
; Sequence 945, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 945
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1505790CD1
US-09-976-594-945

```

Alignment Scores:

```

Pred. No.: 2 7e-08 Length: 289
Score: 171.50 Matches: 55
Percent Similarity: 42.11% Conservative: 33
Best Local Similarity: 26.32% Mismatches: 74
Query Match: 16.33% Indels: 47
DB: 4 Gaps: 8

```

```

US-09-867-753-1 (1-577) x US-09-976-594-945 (1-289)
Qy      2 CGAACAATAGGGGCTCCACCACTAGGGCGTCTGCTCCAGACACCGGTCTTACTGCG 61
Db      30 ProSerGlnGluSerProThrLeuProGluSerSerAlaThrAspSerPyrTyr--- 48
Qy      62 CTGAGTGATACAGGTAAATAAAGCCCAACCTCAGCTGGGGGAGCATCAAGCGCA 121
Db      49 -----SerProThr-----GlyGlyAlaPro 55
Qy      122 GAAGCCATGTGGCCAGAGACTCCAGGCTTCATGAGTAATTAAGACCTTGAGGGCGGT 181
Db      56 HisGlyTyrCysSerProThrSerAlaSerTyrGlyLysAlaLeuAsnProTyrGlnTyr 75
Qy      182 GTGAACACAGAGAGCGCATGAACCGCGATGAGCGCATATCCCGAG----- 229
Db      76 GlnTyrHis-----GlyValAsnGlySerAlaGlySerTyrProAlaLysAlaTyrAla 93

```

```

QY 220 -----GGCGCGCGTGAACCAACGAGCCTCGG 256
DB 94 AspyrSeryrAlaSerSerTyrrHisGlnTyrglyValAlaTyrrAsnArgValProSer 113
QY 257 CAGCAGCCGAGCCCCCGCGAG-----GAGCCGCGCCAGCGCGCCATGAGAGGCTCGG 310
DB 114 AlaThrAsnGlnProGlnIuValThrGluProGlnIuValArgMetValAsnGlyLys 133
QY 311 CAGCCCGAGAACATGAGCAGCACTCGGCGCAGAACTTCACGCTGTTCAGAGTGGAG 370
DB 134 ProLysLysValArgLysProArgThrIleTyrrSerSerPhe-----GlnLeuAla 150
QY 371 GAGCTGAAAGTGTTCGACACACTCAATACCTCGATGCCCGCAGAAAGGAACTT 430
DB 151 AlaLeuGlnArgArgPheGlnIuValThrGlnIuValLeuProGlnIuValArgLys 170
QY 431 GCGGAAACTTGAAGTGTGACTGAAGACAAAGTCCGGGTTGTTAAGATTAAGAGGCC 490
DB 171 AlaAlaSerLeuGlyLeuThrGlnIuValIleTrrPheGlnIuValAsnLysArgSer 190
QY 491 AGATGTAGCCGACATCAGAGAGAAATTAATGCTCCCAATGAACTA----- 535
DB 191 LysIleLysLys-----IleMetLysAsnGlyGluMetProGlnIuHisSer 206
QY 536 -----CGTGTGACCCAGACGACTGT 556
DB 207 ProSerSerSerAspProMetAlaCys 215

```

RESULT 11

US-08-957-351-26
Sequence 26, Application US/08957351
Patent No. 6306586

GENERAL INFORMATION:

APPLICANT: Semina, Elena
APPLICANT: Murray, Jeffrey C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HONG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 271 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-351-26

Alignment Scores:

Pred. No.: 4.13e-08 Length: 271

Score: 169.50 Matches: 38
Percent Similarity: 54.29% Conservative: 19
Best Local Similarity: 36.19% Mismatches: 29
Query Match: 16.14% Indels: 19
DB: 4 Gaps: 3

US-09-867-753-1 (1-577) x US-08-957-351-26 (1-271)

```

QY 197 GGCATGAACCGGATGGGCGGATGATCCCGAGGCGCGGTGAACCAACGAGCCTCGG 256
DB 14 GlyLeuGlnLysAspLysGlyGln-----GlnGlyLysAsnGlnLysP----- 27
QY 257 CAGCAGCCGAGCCCCCGCGAGAGCGCGCCAGCGCGCCATGAGAGGCTCGGAGCCC 316
DB 28 -----ValGlyAlaGlnAspProSerLys 35
QY 317 GAGAACATGACGACCACTGCGGCGCAGAACTTCAGGCTTTCAGAGTGGAGAGCTG 376
DB 36 LysLysArgGln---ArgArgGlnArgThrHisPheThrSerGlnGlnIuVal 54
QY 377 GAAAGTGTTCGACACACTCAATACCTGATGCGCCAGAAAGGAACTTCCCGAA 436
DB 55 GluAlaThrPheGlnArgAsnArgTyrProAspMetSerThrArgLysIuValAlaVal 74
QY 437 AACTTGAAGTGTGACTGAAGACAAAGTCCGGGTTGTTAAGATTAAGAGGCCAGATGT 496
DB 75 TrpThrAsnLeuThrGlnAlaArgValArgValIlePheLysAsnArgArgAlaLysTrp 94
QY 497 AGCCGACATCAGAGA 511
DB 95 ArgLysArgGluArg 99

```

RESULT 12

US-09-636-215-842
Sequence 842, Application US/09636215
Patent No. 6620922

GENERAL INFORMATION:

APPLICANT: Xu, Jianshun
APPLICANT: Dillon, Devin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqul
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Ranger, Gary R.
APPLICANT: Retzer, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yaser A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.4271C17
CURRENT APPLICATION NUMBER: US/09/636,215
NUMBER OF SEQ ID NOS: 852
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 842
LENGTH: 241
TYPE: PRT
ORGANISM: Homo sapiens
US-09-636-215-842

Alignment Scores:

Pred. No.: 4.48e-08 Length: 241
Score: 169.00 Matches: 60
Percent Similarity: 39.52% Conservative: 23
Best Local Similarity: 28.57% Mismatches: 60
Query Match: 16.10% Indels: 67

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 16, 2004, 18:56:12 ; Search time 46 Seconds
(without alignments)
7915.391 Million cell updates/sec

Title: US-09-867-753-1

Perfect score: 1050

Sequence: 1 tccatcaccgagcgcctccag.....tctacatgcgtctgactag 577

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+np.model -DB=XLP
-Q=/cgn2.1/USPTO.spool.p/US09867753/runac.16062004.163635.27793/app_query.faeta.1.775
-DB=SPTRMBL.25 -QFWT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MITEM=0 -MAXLEN=2000000000
-USER=US09867753@CEN1.1.86@runac.16062004.163635.27793 -NCPV=6 -ICPV=3
-NO MAP -LANG=ENGLISH -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL.25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARYS

Result No.	Score	Query Match	Length	DB ID	Description
1	298.5	28.4	406	Q8N693	Q8N693 homo sapien

2	298.5	28.4	406	4	Q7Z6K7	Q7Z6K7 homo sapien
3	212.5	20.2	227	11	Q9E0M5	Q9E0M5 mus musculus
4	212	20.2	286	5	Q8T6T5	Q8T6T5 branchiostro
5	202.5	19.3	317	13	Q8A1U4	Q8A1U4 gallus gall
6	200.5	19.1	391	5	Q8T8C1	Q8T8C1 gryllus bim
7	200	19.0	248	13	Q12952	Q12952 cynops pyrr
8	197	18.8	387	11	Q88933	Q88933 mus musculus
9	197	18.8	387	11	Q922U3	Q922U3 mus musculus
10	196	18.7	314	11	Q54817	Q54817 mus musculus
11	196	18.7	408	5	Q9VPE1	Q9VPE1 drosophila
12	192	18.3	328	13	Q9W7M5	Q9W7M5 brachydanio
13	190	18.1	227	11	Q8BPD6	Q8BPD6 mus musculus
14	190	18.1	429	5	Q7Z0W2	Q7Z0W2 lytechinus
15	190	18.1	640	5	Q8T0M4	Q8T0M4 drosophila
16	189.5	18.0	325	11	Q80YJ6	Q80YJ6 mus musculus
17	188.5	18.0	371	5	Q61282	Q61282 hemisceroc
18	188	17.9	227	11	Q70Z38	Q70Z38 mus musculus
19	187	17.8	257	4	Q86YR1	Q86YR1 homo sapien
20	185.5	17.7	430	5	Q7Z0W3	Q7Z0W3 strongyloce
21	184.5	17.6	226	6	Q9GLL9	Q9GLL9 sus scrofa
22	184.5	17.6	318	5	Q8EB70	Q8EB70 drosophila
23	183.5	17.5	287	5	Q8WNG3	Q8WNG3 patella vul
24	183	17.4	299	13	Q7SZN8	Q7SZN8 brachydanio
25	182.5	17.4	826	5	Q9B130	Q9B130 ciona intes
26	182	17.4	185	13	P79857	P79857 pleurodeles
27	182	17.3	374	5	Q81FX1	Q81FX1 ciona intes
28	181.5	17.3	240	11	Q8R4I3	Q8R4I3 mus musculus
29	180	17.1	185	11	Q8C477	Q8C477 mus musculus
30	180	17.1	299	6	Q8SQ03	Q8SQ03 canis famil
31	180	17.1	299	11	Q9WTQ9	Q9WTQ9 rattus norv
32	180	17.1	329	11	Q9JL7E	Q9JL7E rattus norv
33	179.5	17.1	362	5	Q21836	Q21836 caenorhabdi
34	179	17.0	301	5	Q46170	Q46170 tribolium c
35	177.5	16.9	276	13	Q98TG7	Q98TG7 gallus gall
36	177	16.9	387	13	Q93582	Q93582 gallus gall
37	176.5	16.8	394	5	Q9XYQ3	Q9XYQ3 anopheles g
38	176.5	16.8	503	11	Q9ES16	Q9ES16 mus musculus
39	176	16.8	214	13	Q8UVD3	Q8UVD3 gallus gall
40	176	16.8	284	4	Q81VZ2	Q81VZ2 homo sapien
41	176	16.8	370	5	Q25411	Q25411 lineus sang
42	176	16.8	403	4	Q86UQ3	Q86UQ3 homo sapien
43	176	16.8	407	4	Q86UQ2	Q86UQ2 homo sapien
44	176	16.8	479	11	Q9CX16	Q9CX16 mus musculus
45	176	16.8	479	11	Q8BRE7	Q8BRE7 mus musculus

ALIGNMENTS

RESULT 1
Q8N693 PRELIMINARY; PRT; 406 AA.
ID Q8N693;
AC Q8N693;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE ESX1L.
GN ESX1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pooled germ cell tumors;
RX MEDLINE=2169184; PubMed=11374906;
RA Fohn L.E., Behringer R.R.;
RT "ESX1L, a novel X chromosome-linked human homeobox gene expressed in the placenta and testis";
RT Genomics 74:105-108(2001).
RL
RN
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pooled germ cell tumors;
RA Fohn L.E., Behringer R.R.;

[illegible]

DE 01-OCT-2003 (TrEMBLrel.25, last annotation update)
 DT Extrachromosomal, spermatogenesis, homeobox 1-like.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22389257; PubMed=12477932;
 RA Strausberg R.U., Feingold E.A., Grouse L.H., Deng J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 Rana S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mollari S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek U.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shcherbenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,
 Krzywinski M.I., Skalska U., Smailus D.E., Schermer A., Schein J.E.,
 Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC053599; AAH53599.1; -;
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 406 AA; 44311 MW; D8B9C18FB923BB9 CRC64;

Alignment Scores:
 Pred. No.: 1,166-18 Length: 406
 Score: 298.50 Matches: 71
 Percent Similarity: 49.21% Conservative: 22
 Best Local Similarity: 37.57% Mismatches: 39
 Query Match: 28.43% Indels: 57
 DB: 4 Gaps: 6

US-09-867-753-1 (1-577) x Q7Z6K7 (1-406)

QY CCTGAGGGCGGGGTGATCAACAGAGAAACGGCATGAACCCGGATGCGGCATGATCCCC--- 226
 Db 170
 51 ProGluTYRGIYThrGIuAlaGluAshaMhaenValIGIYThrGIuGIYser---ValProSer 69
 QY 227 -----GAGGCGCGCGGTGGAACAGAGACCTCGGCGAGACCGCGAGCCC 271
 Db 227
 70 AspaapGlnaPaPaRGIuGIYglYglYglYgHISgluProGIuGIuGlnGlnGluGIuPro 89
 QY 272 CCG----- 274
 Db 272
 90 ProLeuThrLYsProGIuGlnGlnGlnGluGIuProProLeuLeuGluLeuLYsGlnGlu 109
 QY 275 CCGAGAGAGACCGCGCCAGGCGCGCATGAGGGTCCGAGACC----- 316
 Db 275
 110 GlnGluGIuProProGIuGlnThrThrValGluGluGIYProGIuProAlaGluGIYProGIuThr 129
 QY 317 ---GAGAAATGACGACCA-----CGAACTCGGCGCAAGAAATTACAGCTGTTG 361
 Db 317
 130 AlaGIuGIYProGIuProProGIuAlaRglYAsaRGAArgARgThrAlaPheThrGlnPhe 149
 QY 362 CAGGTGAGAGAGAGCTGGAAGTGTTCGAGACACATCAATACCGTAGTGGCCACAGA 421
 Db 362
 150 GlnLeuGlnGluLeuGlnAshaPhePaPaRGIuSerGIuThrYProAspValValAlaIaArg 169

Oy	422	AGGGAATTGGCCGAAATCTTACGTGTGACTGAGAACAAGTCGGGTTTGCTTAAGAT	481
Dd	170	GIUATgLeuaLaLaLaTgLeuaLnLeuThrGluAspValAlaIValITrpheGlnAsn	189
Oy	482	AAAAAGGCAGCATGTAGCGGCATCAGAGAATAATTATCTCGCCAATGAACTACGTCT	541
Dd	190	ArgatgAlaLetytRlykxghangInAkgValLeuMetLeuakxgnhThralaTrha	209
Oy	541	-----	541
Dd	210	AspleuaLaHIsProlEuaBpWetPhelEugLyLaTyrtYtRlaLaLaProalaLeu	229
Oy	542	GACCAGACGACTGTGTCTACATCGC	568
Dd	230	AsPrroalaleuCysValHisleuVal	238
 RESULT 3 Q9EBM5 PRELIMINARY; PRT; 227 AA.			
AC	Q9EBM5;		
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)		
DE	Homeobox protein GPBOX.		
GN	PSX2 OR I6002600IRIK OR GPBOX.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
NCBI_TaxID=10090;			
RN	(1)		
RP	SEQUENCE FROM N.A.		
RA	Takasaki N., McInaac R., Dean J.;		
RT	"Gbox, a novel homeobox gene preferentially expressed in female germ		
RT	cells at the onset of sexual dimorphism in mice.";		
RL	Submitted (NOV-1999) to the EMBL/Genbank/DBD databases.		
DR	EMBL; AF201698; AAC36768.1; -		
DR	MCD; MGI:1890128; Pax2.		
DR	GO: GO:0005634; C:nucleus; IEA.		
DR	GO: GO:0003700; F:transcription factor activity; IEA.		
DR	GO: GO:0006355; F:regulation of transcription, DNA-dependent; IEA.		
DR	InterPro; IPRO01356; Homeobox.		
DR	Pfam; PF00046; homeobox; 1.		
DR	Prodom; PD000010; Homeobox; 1.		
DR	SMART; SMO0389; HOX; 1.		
DR	PROSITE; PS0071; HOMEBOX 2; 1.		
SO	SEQUENCE 227 AA; 25112 MW; D39CF6795AF68ZDC CRC64;		
 Alignment Scores:			
	Pred. NO.:	7.59e-11	Length: 227
	Score:	212.50	Matches: 61
	Percent Similarity:	45.61%	Conservative: 17
	Best Local Similarity:	35.67%	Mismatches: 60
	Query Match:	20.24%	Indels: 33
		Gaps: 6	
 US-09-867-753-1 (1-577) x Q9EBM5 (1-227)			
Oy	104	GGGCGACATCAAGCGCACAAAGGCCATTTGGCCCAAGACTCCAGGCTCATGGTAA	163
Dd	54	GIyGLyLeuaBpGInGLyGLyLaIngLyLaIValAlaGLyGLyGLInLaIngIn	73
Oy	164	ATGAACCT-----GAGGGCGGTGTGAACAACGAGAACGGCATG	202
Dd	74	GIUGLuProlaHIsProlEuaSerProaLIngInLualThrGLyGLyGLInGLyGLIn	93
Oy	203	AACCGCATGGCGGCATG-----	220
Dd	94	AsnLyGSGLuGLyGLuWetGluGLyArGHIsALaGLyAspOLyAlasergLYPProGLuAsp	113
Oy	221	-----ATCCCGGAGGGCGCGGTGAAC-----CAGAGGCTCGGCGAGCGCGAG	268
Dd	114	AspaenILegInGLuGLInGLyGLyGLuAanILiaspGLInGLInProBrogInGLuAla	133

Oy		269	CCCCCGCCGAGAGCGGGCCAGGCCGCACATGAGCGGTTCGCGAGCCCGAAGC-----	3222
Dd	-	134	AAlaIlePProGlu-----GlyMetArgAsnProGlnAlaGlyAsnTyrLeu	148
Oy		323	ATGCAGCCCAAGACTCGCGCGCACGAAGTTCAACGCTGTTGCAGGTGAGAGAGCTGAAAGT	3822
Dd	-	149	AlAhIEGLInArGrThrArgTrgThrArgPhenThrHisSerGlnLeuArgAspLeuGlUArg	148
Oy		383	GTTTTCCGACACACTCATACCTTGATGTGCCCSACAAGAAGCAATTGCCGAAAACCTTA	4442
Dd	-	169	LenPhEgLIglUsArNArgPheProSerLeuArgValAlaGuaArgAspLeuAlaArgTyrMet	188
Oy		443	GGGTACTACGAGAACAAAGATCGGGGTTTGATTAAGATAAAGGCGCACAGTAGAGCGCA	5022
Dd	-	189	GIYAlAspGILusErAspValGlnIdnUrPheYsMeTarGalRAlaLeuPhaArg	208
Oy		503	CATCAGAGAGAAATTAACTGCTCGCCAAATGACTA	535
Dd	-	209	HisSerArg---LeuMetMetPheCysGluLeu	218
 RESULT 4				
ID	Q87615	PRELIMINARY;	PRT;	286 AA.
AC	Q87615;			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)			
DS	Paired-like homeodomain transcription factor Shox.			
GN	SBOX.			
OS	Branchiostoma floridae (Florida lancelet) (Amphioxus).			
OC	Eukaryote; Metazoa; Chordata; Cephalochordata; Branchiostomidae;			
CC	Branchiostoma.			
OX	NCBI_TaxID=7739;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Jackman W.R., Jr., Kimmel C.B.;			
RT	"Coincident iterated gene expression in the amphioxus neural tube."			
RL	Submitted (JAN-2002) to the EMBL/Genbank/DDBJ databases.			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).			
DR	EMBL; AF465939; MAF83210.1; ..			
DR	GO; GO:0005634; C:nucleus; IEA.			
DR	GO; GO:0003700; F:transcription factor activity; IEA.			
DR	GO; GO:0007275; P:development; IEA.			
DR	GO:GO:0006355; R:regulation of transcription, DNA-dependent; IEA.			
DR	InterPro; IPR001356; Homeobox			
DR	InterPro; IPR003654; Homeo_OAR.			
DR	InterPro; IPR000047; HTH_LambdaRepressr.			
DR	InterPro; IPR007104; Paired_homeo.			
DR	pFam; PF00046; homeobox; 1.			
DR	pFam; PF03826; OAR; 1.			
DR	PRINTS; PR00031; HTHREPRESSR..			
DR	ProDom; PD000010; Homeobox; 1.			
DR	SMART; SM00389; HOX; 1.			
DR	PROSITE; PS00027; HOMEBOX_1; 1.			
DR	PROSITE; PSS0071; HOMEBOX_2; 1.			
KW	Homeobox; DNA-binding; Nuclear protein.			
SO	SEQUENCE 286 AA; 31551 MW; 6BF50B62D92958C1 CRC64;			
 Alignment Scores:				
	Pred. No.:	8.6e-11	Length:	286
	Score:	212.00	Matches:	52
	Percent Similarity:	48.10%	Conservative:	24
	Best Local Similarity:	32.91%	Mismatches:	48
	Query Match:	20.19%	Indels:	34
	DB:	5	Gaps:	5
 US-09-867-753-1 (1-577) x Q87615 (1-286)				
Oy		104	GGGGCAGCATCAAGCCCAAGAGCCAT-----GTTGGCCAAGAGCTCCAGGCTCATG	157
Dd	-	46	GIYAlaGILusErAspValAspGILusAspLeuCysPrOvalGIYSePProGILeu---	64

```

QY 158 GGTAATATGAAACCTTGAGGGCGGTGTGAACACGAGAACGGCATGAACCGCGATGGCGG 217
Db ----- 64
QY 218 ATGATCCCGGAGGGC---GGCGGTGGAACACGAGAGCT-----CGCAGCAGCCG 265
Db -----ProvaldylleuylglyylglylvaanserProglulubnlysaerPro 82
QY 266 CAGCCCCCGGAGGAGCGCCGCGCCATGAGAGGGGTCCGACGAGCCCGAGAACATG 325
Db 83 SerlyrProgllyglululyslsaspaerleuuserProanserlyshapgllylsleu 102
QY 326 CAGCCACGAATCTCGCGCAGAACTTACGCTGTTCAGAGTGAGGAGAGCTGGAATGTT 385
Db 103 LysGlnAlarGserArgrThraenPheThrleuGlnGlnleuGlnGlnleuGlnArgleu 122
QY 386 TTCGACACACCTGAATACCTGTATGTGCCACAGAAGGAACTTCCGAAACTTAGGT 445
Db 123 PheAspGlnuthrhislyrProaspaAlaPheMetArgrGlnleuuserGlnArgleu 142
QY 446 GTGACTGAAGACAAAGTGGCGGTTTGTGTTAAGATAAAGGCGCATGTAGCGACAT 505
Db 143 LeuSerGlnAlaArgValGlnValTrrPheGlnAsnArgAlaValCysArglyslin 162
QY 506 CAGAGAGAAATTA-----ATGCTGCCAATGAACTAGCTGTGAC 544
Db 163 GluAsnGlnleuGlnlysllysllysllysllysllysllysllysllysllysl 180

RESULT 5
Q8AYJ4 PRELIMINARY; PRT; 317 AA.
AC Q8AYJ4;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Homeobox transcription factor RAX1.
GN RAX1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22291598; PubMed=12403708;
RA Chen C.M., Cepko C.L.;
RT "The chicken Rax1 gene plays a role in the initiation of photoreceptor
RT differentiation."
RL Development 129:5363-5375(2002).
DR EMBL; AF420600; AAN32718.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003654; Homeo_OAR.
DR InterPro; IPR007104; Paired_homeo.
DR Pfam; PF00046; homeobox; 1.
DR Pfam; PF03826; OAR; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
SQ SEQUENCE 317 AA; 34022 MW; E279B75ECD45FC9D CRC64;

```

```

US-09-867-753-1 (1-577) x Q8AYJ4 (1-317)
QY 119 GCAGAGGCCATGTTGGCCAGAGAGCTCCAGGCTCATG-----GGTAATATGAAC 169
Db 11 AlaglnGlyAlaPheSerleuSerAlaPheAlaAlaArgSerProgllyslanProser 30
QY 170 CCTGAGGGCGGTGTGAACACGAGAACCGCATGAACCGGATGGCGCATGATC----- 223
Db 31 ArgLeuHisSerlleuAlaIleleuGlyPheThrlyAspAspIlyleuGlnPro 50
QY 224 -----CCGAGGGCGCGGTGGAACACGAGAG----- 250
Db 51 PheGlnProAspGlyAlaIleGlySerAlaIleGlyAlaIleAspIlyslarglyProArg 70
QY 251 -----CTCGAGAGCGCGCGCCCGCGCGAGAG----- 283
Db 71 HisCysleuProIySgIyProAlaGlnProProAlaGlnHisGlnIlyArgPheGln 90
QY 284 -----CCGCGCCAGCGCGCATGAGGTT 307
Db 91 GluProTyrcysrProgllyserAlaSerProGlnleuProAlaGlyAspIlyGlyAspIly 110
QY 308 CCGCAGCCCGGAACAATGAGCCA-----CGAATCGCGCAGCAAGTTCAAG 355
Db 111 LysProSerAspGlnGlnGlnProIySlyshsArArgAsnArgThrPheThr 130
QY 356 CTGTTGACAGTGGAGAGCGTGAAGTGTTCGACACACATCAATACCTGATGTGCC 415
Db 131 ThrTyGlnleuHisGlnleuGlnArgAlaPheGlnIySerHisIytrProAspValTyx 150
QY 416 ACAGAGAGGAATCTGCCGAATACTTAGGTGTGACTGAAGACAAAGTGGGTTGTTT 475
Db 151 SerArgGlnGlnIleuAlaMetlyslValaenleuProGlnValArgValGlnValTrrPhe 170
QY 476 AAGAATAAAGGCGCAGATGTAGCGACATCAGAGAGATTAATGCTGCCAATGAACTA 535
Db 171 GlnAsnArgArgAlaIyetrPargGlnGlnIyLeuGlnValSerSerMetlyslleu 190
QY 536 CGTGCTGACCCA 547
Db 191 GlnAspSerPro 194

RESULT 6
Q8T8C1 PRELIMINARY; PRT; 391 AA.
AC Q8T8C1;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Aristaless protein.
GN GBAI.
OS Gryllus bimaculatus (Two-spotted cricket).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Phryganea;
OC Neoptera; Orthopteroidea; Orthoptera; Ensifera; Gryllidae; Gryllinae;
OC Gryllus.
OX NCBI_TaxID=6999;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyawaki K., Inoue Y., Mito T., Matsushima K., Shimmyo Y., Ohuchi H.,
RA Noji S.;
RT "Expression patterns of aristaless in appendage morphogenesis of the
RT hemimetabola, Gryllus bimaculatus (the cricket).";
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AB071147; BAB85815.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003654; Homeo_OAR.
DR InterPro; IPR00047; HTH_lamdrepressr.

```

[illegible]

CR TISSUE=Embryo;
 RA Sone K., Takeshima K., Takahashi T., Takabatake T.;
 RL Dev. Genes Evol. 207:147-155(1997).
 RX [2].
 RP SEQUENCE OF 129-175 FROM N.A.
 RA MEDLINE=9613634; PubMed=8573168;
 RX Takabatake T., Takahashi T.C., Inoue K., Ogawa M., Takeshima K.;
 RT "activation of two Cypnos genes, fork head and sonic hedgehog, in
 animal cap explants";
 RL Biochem Biophys Res Commun. 218:395-401(1996).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AB001938; BAA2405.1; -.
 DR EMBL; D64022; BAA10897.1; -.
 DR HSSP; P06601; 1FTU.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006555; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR007104; Paired_homeo.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox_1.
 DR ProDom; PD000010; Homeobox_1.
 DR SMART; SMO0389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00027; HOMEBOX_2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 QO SEQUENCE 248 AA; 27667 MW; 98BDA7D61DB59D6 CRC64;

[illegible]

```

US-09-867-753-1 (1-577) x 012952 (1-248)

Oy      2  CCAACATCAGGCGGCTTCAGGCATGGCGCGCTTGCTGTCGTC---CAAGACACCGTGTTCAC  58
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      58  ProthAlaGlyLeuProthRheLeuAlaGlySerValGlyGlyTyrAsnGlyGlyTyrTyr  77
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy      59  TGCGTAGGTGTATACCGAGGTAAATTAAGCCCAACACTGACGCTGGGGGAGCATCAAGC  118
           |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      78  -----TyrGlyLeuLeuHisLeuGlnPro-----  85
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy      119 GCAGAAAGCCATGTTGGCCCAAGAGACTCCAGGCTCATGGTAAATATGACCTGAGGAGC  178
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      86  -----ProGlyAlaProGlyGlyCysGlyAlaIleGlnProLeuGly  99
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy      179 GGTGTGAACCAAGAAACGGCATGAACCGGATGGCGGATATGCCCGAGGAGCGGCGGT  238
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      100 Ala-----GlnGlnCysSerCysValProAlaProSerSer  111
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy      239 GGAATCCAGAGAGCTTGCGGACGACGCCGAGCCCCCGCGAGAGAGCGGAGCCAGCGGC  296
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      112 Gly-----TyrAspGlySerSerVal  119
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy      299 ATGAGAGGTCCTCCGAGCCGAGAAACATGCAGCCA-----  332
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      120 LeuMetSerProMetProHisGlnMetMetProTyrMetAsnValGlyThrLeuSerArg  133
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy      332 -----CGAACTCGGCGGACGAGAG  345
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      140 SerGlnLeuGlnLeuLeuAsnGlnLeuHisCysArgArgGlyValArgArgHisArgThrIle  155
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy      350 TTCACGCTGTTCGAGGTGAGAGACTCGAAAGCTGTTTCCAGACACTCAATATCCCTGAT  408
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      160 PheThrAspGlnLeuGlnIleValLeuGlnHisIleLeuPheGlnGlnIleThrValTyrProAsp  179
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy      410 GTGCCCAAGAAAGGAGATCCGCGGAAATCTTAGGTGAGCTGAAAGCAAGATGCGGGTT  465
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      180 ValGlyThrArgGlnIleLeuAlaArgValHisIleMetArgGlnGlnValGlnValGlnVal  196
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

QY 470 TGGTTTAAGATATAAGGCCAGATGTAGCCGACATCAGAGA 511
 DB 200 TTPhelyaAsnArgAlaGlySTpArgArgGlnIlyeArg 213
 RESULT 8
 ID 088933 PRELIMINARY; PRT; 382 AA.
 AC 088933;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Homeobox protein SPXL.
 DE ESX1 OR EPX.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97398449; PubMed=9256347;
 RA Brantford W.W., Zhao G.Q., Valerius M.T., Weinstein M.,
 RA Birkmeier E.H., Rowe L.B., Potter S.S.;
 RT "Spxl, a novel X-linked homeobox gene expressed during
 RT spermatogenesis.";
 RT Mech. Dev. 65:87-98 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97392770; PubMed=9245514;
 RA Li Y., Lemaire P., Behringer R.R.;
 RT "Esxl, a novel X chromosome-linked homeobox gene expressed in mouse
 RT extremityonic tissues and male germ cells.";
 RT dev. Biol. 188:85-95 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Brantford W.W., Potter S.S.;
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RA Yan Y.-T., Yang L., Sciavolino P.J., Wang H., Chan D.C.,
 RA Abate-Shen C., Shen M.M.;
 RT "Bpx: a novel paired-like homeobox gene expressed in the chorion and
 RT placenta.";
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AF085715; AAC35366.1; -;
 DR EMBL; AF017735; AAD01622.1; -;
 DR HSSP; P06601; IFUL.
 DR TRANSFAC; T03478; -;
 DR MGD; MGI:1096388; Esxl.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR000047; HTH_LambdaRepressr.
 DR InterPro; IPR007104; Paired_homeo.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR PRODOM; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 382 AA; 43540 MW; EC92301A84DA6175 CRC64;
 Alignment Scores:
 Pred. No.: 2.05e-09 Length: 382
 Score: 197.00 Matches: 59
 Percent Similarity: 44.51% Conservative: 18
 Best Local Similarity: 34.10% Mismatches: 72
 Query Match: 18.76% Indels: 24
 DB: 11 Gaps: 5

US-09-867-753-1 (1-577) x 088933 (1-382)
 QY 71 TACCAGGTAAATAATAGCCACACCTCAGCTGGGGGAGCATCAAGCGCAGAGGCCAT 130
 DB 84 TyrgIngluProGluGlyPheGluProSerArgGlyGlnAlaAlaPro----- 100
 QY 131 GTTGGCCAAAGAGCTCCAGGCTCATGGGTAAATATGAACCTGAGAGGCGGTGAACAC 190
 DB 101 VALAAGLUALProGlnAlaTrpAsnGlyAsnGlnGlyGlyPheLeuGlu 120
 QY 191 GAGAACCGCATACCGCGATGGCGG----- 217
 DB 121 SerAsnAlaGlnLeuGlyGlnAlaAspAlaAlaProValArgGlnSerLeuMetArgPro 140
 QY 218 ---ATGATCCCGAGGGCGCGGTGMAACAG-----GAGCTCGGACAGCAG 262
 DB 141 LeuMetGlnProValAlaGlnSerProGlnProLeuProAlaAsnProLeuGlnAla 160
 QY 263 CCGCAGCCCCG-----CCGAGAGCGCGGCCAGCGGCATGAGGGTCCG 310
 DB 161 ProGlnGlnProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 180
 QY 311 CAGCCCGAAGATGACAGCCAGACTCGCGCAGAGTTCACGCTGTGGAGGAGGAG 370
 DB 181 GlnGlnGlu---ProLysProArgArgTyrArgIleCysPheTrpProIleGlnGluGln 199
 QY 371 GAGCTGAAGAGTGTTCGCCACACTCAATACCTGATGTGCCACAAAGAGGAACCT 430
 DB 200 GlnLeuGlnAlaPhePheGlnArgValGlnTyrProAspLeuPheAlaArgValGlnLeu 219
 QY 431 GCCGAAACTTGGGTGACTGAAGACAAAGTCCGGCTTGGTTTAAGATTAAGAGGCC 490
 DB 220 AlaArgArgLeuGlyLeuProGluProArgValGlnValTrpPheGlnAsnArgArgAla 239
 QY 491 AGATGTAGGCGCATCAGAGAGAAATTAATGCTCGGCAAT 529
 DB 240 LysTrpArgArgLeuArgArgAlaGlnAlaPheArgAsn 252
 RESULT 9
 ID 092203 PRELIMINARY; PRT; 387 AA.
 AC 092203;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Homeodomain protein EPX (Fragment).
 GN ESX1 OR EPX.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=20079279; PubMed=10611245;
 RA Yan Y.-T., Stein S.M., Ding J., Shen M.M., Abate-Shen C.;
 RT "A Novel PF/PN Motif Inhibits Nuclear Localization and DNA Binding
 RT Activity of the ESX1 Homeoprotein.";
 RL Mol. Cell. Biol. 20:661-671 (2000).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AF017734; AAD01621.1; -;
 DR HSSP; P06601; IFUL.
 DR MGD; MGI:1096388; Esxl.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR000047; HTH_LambdaRepressr.
 DR InterPro; IPR007104; Paired_homeo.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR PRODOM; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS0071; HOMEBOX 2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 FT NON TER 1
 SQ SEQUENCE 387 AA; 44170 MW; 0D6A6E4C36CF9CD3 CRC64;

Alignment Scores:

Pred. No.: 2, 06e-09 Length: 387
 Score: 197.00 Matches: 59
 Percent Similarity: 44.51% Conservative: 18
 Best Local Similarity: 34.10% Mismatches: 72
 Query Match: 18.76% Indels: 24
 DB: 11 Gaps: 5

US-09-867-753-1 (1-577) x Q9Z2U3 (1-387)

QY 71 TACGAGTAAATAATAGCCCACTCAGTGGGGGAGCATCAAGCCGAGAGGCCAT 130
 DB 89 TyrgIngluProgluGlyPhegluProserArglyGluAlaAlaPro----- 105
 QY 131 GTTGGCCAAAGAGCTCCAGGCTCATGGTATATGAAACCTGAGGGCGGTGTAACAC 190
 DB 106 ValAlaGluAlaProGlnAlaTTrpAnnglyAsnGluAsnLeuGlyGlyPheLeuGlu 125
 QY 191 GAGAACGGCATGACCGCATGGCGGC----- 217
 DB 126 SerAenAlaGlnLeuGlyGluAlaAspAlaAlaProValArgGlnSerLeuMetArgPro 145
 QY 218 --ATGATCCCGAGGGCGCGGTGMAACAG-----GAGCTCGGCGAGCAG 262
 DB 146 LeuMetGlnProValAlaGlnSerSerProGlnProLeuProAlaAsnProLeuGlnAla 165
 QY 263 CGGAGCCCGCCG-----CCGAGAGCGCGCCGAGCGCCCATGAGGGTCCG 310
 DB 166 ProGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 185
 QY 311 CAGCCCGAAGACATGACGACGAACTCGCGCAGCAAGATTGCTGTGAGAGTGAG 370
 DB 166 GlnGlnGln---ProlyeProArgArglyTyArgIleCyPheThrProIleGlnLeuGln 204
 QY 371 GAGCTGAAAGTGTTCGACACACTCAATACCTGATGTGCCCAAGAGGAACTT 430
 DB 205 GluLeuGlnAlaPhePheGlnArgValGlnTyTrpAspLeuPheAlaArgValGluLeu 224
 QY 431 GCCGAAACTTGTGTGTGCTGACGACGAAAGTGGGCTTGTGTTAGAAATAAGGGCC 490
 DB 225 AlaArgArgLeuGlyLeuProGlnProArgValGlnValTTrpPheGlnAsnArgArgAla 244
 QY 491 AGATGTAGCGCATCAGAGAGAAATTAATGCTCGCCAAAT 529
 DB 245 LysTTrpArgArgLeuArgArgAlaGlnAlaPheArgAsn 257
 RESULT 10
 054817 PRELIMINARY; PRT; 314 AA.
 054817
 AC 054817
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)
 DE Paired-like homeodomain containing protein.
 GN ESX1 OR ETK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss;
 RA MEDLINE=97932770; PubMed=9245514;
 RA Li Y., Lemaire P., Behringer R.R.;
 RT "Esx1, a novel X chromosome-linked homeobox gene expressed in mouse
 RT extraembryonic tissues and male germ cells.";
 RT Dev. Biol. 188:85-95(1997).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AF004211; AAB94670.1; -.
 DR HSSP; P06601; 1FJL.
 DR TRANSFAC; T03474; -.

DR MGD; MGI:1096388; Esx1.
 DR GO; GO:0005634; C:nucleus, IEA.
 DR GO; GO:0003700; F:transcription factor activity, IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR01356; Homeobox.
 DR InterPro; IPR000047; HTH_lambdarepressr.
 DR InterPro; IPR007104; Paired homeo.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS0071; HOMEBOX 2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 314 AA; 35969 MW; 72568F81DA4AA246 CRC64;

Alignment Scores:

Pred. No.: 2, 49e-09 Length: 314
 Score: 196.00 Matches: 59
 Percent Similarity: 44.51% Conservative: 18
 Best Local Similarity: 34.10% Mismatches: 72
 Query Match: 18.67% Indels: 24
 DB: 11 Gaps: 5

US-09-867-753-1 (1-577) x 054817 (1-314)

QY 71 TACGAGTAAATAATAGCCCACTCAGTGGGGGAGCATCAAGCCGAGAGGCCAT 130
 DB 16 TyrgIngluProgluGlyPhegluProserArglyGluAlaAlaPro----- 32
 QY 131 GTTGGCCAAAGAGCTCCAGGCTCATGGTATATGAAACCTGAGGGCGGTGTAACAC 190
 DB 33 ValAlaGluAlaProGlnAlaTTrpAnnglyAsnGluAsnLeuGlyGlyPheLeuGlu 52
 QY 191 GAGAACGGCATGACCGCATGGCGGC----- 217
 DB 53 AlaAenAlaGlnLeuGlyGluAlaAspAlaAlaProValArgGlnSerLeuMetArgPro 72
 QY 218 --ATGATCCCGAGGGCGCGGTGMAACAG-----GAGCTCGGCGAGCAG 262
 DB 73 LeuMetGlnProValAlaGlnSerSerProGlnProLeuProAlaAsnProLeuGlnAla 92
 QY 263 CGGAGCCCGCCG-----CCGAGAGCGCGCCGAGCGCCCATGAGGGTCCG 310
 DB 93 ProGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 112
 QY 311 CAGCCCGAAGACATGACGACGAACTCGCGCAGCAAGATTGCTGTGAGAGTGAG 370
 DB 113 GlnGlnGln---ProlyeProArgArglyTyArgIleCyPheThrProIleGlnLeuGln 131
 QY 371 GAGCTGAAAGTGTTCGACACACTCAATACCTGATGTGCCCAAGAGGAACTT 430
 DB 132 GluLeuGlnAlaPhePheGlnArgValGlnTyTrpAspLeuPheAlaArgValGluLeu 151
 QY 431 GCCGAAACTTGTGTGTGCTGACGACGAAAGTGGGCTTGTGTTAGAAATAAGGGCC 490
 DB 152 AlaArgArgLeuGlyLeuProGlnProArgValGlnValTTrpPheGlnAsnArgArgAla 171
 QY 491 AGATGTAGCGCATCAGAGAGAAATTAATGCTCGCCAAAT 529
 DB 172 LysTTrpArgArgLeuArgArgAlaGlnAlaPheArgAsn 184
 RESULT 11
 09VPE1 PRELIMINARY; PRT; 408 AA.
 AC 09VPE1
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, last annotation update)

AL gene product (RE68460p).
 GN AL OR CG3935.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-U., Andrews-Planck C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Casley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
 RA Jitali M., Kalish F., Kapen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maitre B., McIntosh T.C., McLeod M.P., McPherson A.,
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pitsman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Svirer E., Spralling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 (12)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Fatian D., Frise E.,
 RA George R., Gonzalez M., Guatari H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacle J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celinker S.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBS databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AEO003589; AAF51505.1; -;
 DR EMBL; AY121696; AAM52023.1; -;
 DR PIR; A40685; A40685.
 DR HSSP; P06601; 1FUT.
 DR FlyBase; FBgn0000061; al.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0007275; P:development; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR003654; Homeo_OAR.
 DR InterPro; IPR000047; HTH_LambdaRepressr.
 DR InterPro; IPR007104; Paired_homeo.

DR Pfam; PF00046; homeobox; 1.
 DR Pfam; PF03826; OAR; 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR Prodom; PD000010; Homeobox; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 DR PROSITE; PS00803; OAR; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 408 AA; 43657 MW; 7641316AA77BD8D CRC64;
 Alignment Scores:

	Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	2.55e-09	408	47	47	17	25	
Percent Similarity:	55.17%						
Best Local Similarity:	40.52%						
Query Match:	18.67%						
DB:	5						3

US-09-867-753-1 (1-577) x Q9VPP1 (1-408)
 QY 244 CGAGGAGCCTCGGACGAGCCGAGCCCGCCG-----GGA 279
 Db 31 ProGlySerSerAlaAlaSerAlaGlyAlaAlaLeuThValSerMetSerValSerGly 50
 QY 280 GGAGCCGCGCCGAGCGCGCCATGAGAGGTCCGCA----- 312
 Db 51 GlyAlaProSerGlyAlaSerGlyAlaSerGlyGlyThrAsnSerProValSerAspGly 70
 QY 313 -----GCCGAGACATGCAAGCCACCAACTCGG-----CGCAC 345
 Db 71 AsnSerAspCysGluAlaAspGlyIuTyraIa-ProlysArgLysGlnArgArgTyArgTh 90
 QY 346 GAAGTTCAGCGCTGTCGAGGAGGAGCTGGAAGTGTTCCTCCGACACATCAATACCC 405
 Db 90 rThrPheThrSerPheGlnIleuGluIleuGluIleuGluAlaPheSerArgThHisTyArg 110
 QY 406 TGAATGCCACACAGAGGAAGAACTTCCGAAACTTAGGTGACTGAAGACAAAGTGGC 465
 Db 110 aspValPheThrArgGluGluIleuAlaMetLysIleGlyLeuThGluAlaArgIleG 130
 QY 466 GGTTCGTTTAAGATATAAAGGCGGAGATGTAGGCGACATCAGAGA 511
 Db 130 nAlaTrpPheGlnAsnArgArgAlaLysTrpArgLysGlnGluLys 145
 RESULT 12
 Q9V7M5 PRELIMINARY; PRT; 328 AA.
 AC Q9V7M5; 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Orthopedia protein.
 GN OTP OR OTP.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NCBI_Taxid=7955;
 (1)
 RP SEQUENCE FROM N.A.
 RA Del Giacco L., Di Benedetto B., Duga S., Diani S., Cotelletti F.;
 RT "Isolation of the mRNA encoding Otp (Orthopedia) in the zebrafish,
 RT Danio rerio.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBS databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AF071496; AAD42021.1; -;
 DR HSSP; P06601; 1FUT.
 DR ZFIN; ZDB-GENE-990708-7; otp.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0007275; P:development; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR003654; Homeo_OAR.
 DR InterPro; IPR000047; HTH_lambdarepress.
 DR InterPro; IPR007104; Paired_homeo.
 DR Pfam; PF00046; homeobox_1.
 DR Pfam; PF03826; OAR; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS0071; HOMEBOX_2; 1.
 DR PROSITE; PS00803; OAR; 1.
 DR DNA-binding; Homeobox; Nuclear protein.
 KW SEQUENCE 328 AA; 35488 MW; 7CCAC8489011C3EC CRC64;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
5	79e-09	328	60
192	0.00	60	22
42	0.05	22	65
30	0.77	65	48
18	2.94	48	8

US-09-867-753-1 (1-577) x Q9W7M5 (1-328)

QY 2 CCAACATCAGG-----GCTCCAGCATGCGCGTTCGCTCCAC 43
 DB 18 ProhomerGlyLeuGlnLeuH1STPMetLysAspAlaAlaLeuLeuValHis 37
 QY 44 GACACCGTGTCTACTCCTGAGTATACAGGTAAATAAGCCCACTCAGCTG 103
 DB 38 ArgGluAlaLeuLys-----ArgLeu 45
 QY 104 GGGGAGCATCAAGCGCAGAGCCATGTGTC----- 136
 DB 46 GlyGlyThrAspSer-----GlyHisProGlyAspLeuThrSerAlaThrGluThrVal 63
 QY 137 CAAGGAGCTCCAGGCTCATGGGTATATGAACCTGAGGGCGGTGGAACCAAGAG 196
 DB 64 GluGlyThrThrLeuLeuProGlyGluGluLeuSerHisGlyLysSerHis--ProHis 82
 QY 197 GGCATGAAACCGGATGCGCGATGATCCCGAGGGCGCGGTGGAACCAAGAGCCTCG 256
 DB 83 GlyMetGln-----GlnValAsnAlaLysAspGlnGluLysGln 95
 QY 257 CAGCAGCCGAGCCCGCCGAGAGCGCGCCAGCGCCATGAGGCTCCGAGCCC 316
 DB 96 GlnGln-----GlnAsnSerAsnGlnThrGlyGlyGlnGln 107
 QY 317 GAGAACATCAGCAGCAAGCTCGGCGAGAGTTGACGCTGTTCGAGTGAAGAGCTG 376
 DB 108 AsnGlnGlnLysGlnLysArgHisArgHisArgHisProAlaGlnLeuMetGlnLeu 127
 QY 377 GAAAGTGTTCGACACACTCATACCTGATGTGCCCAAGAGAGGAATCTGCGAA 436
 DB 128 GluArgSerPheAlaLysThrHisTyProAspPhePheMetArgGluGluLeuAlaLeu 147
 QY 437 AACTTGGGTGACTGAAGCAAGAGCGGCTTGTGTTAAGATTAAGAGCGGCAATGT 496
 DB 148 ArgGlyLeuThrThrGlnSerArgValGlnValTyrPheGlnAsnArgArgAlaLysTyr 167
 QY 497 AGGCGCATCAGAGAAATTAATGCTCGGCAATGAAGTACGTCT 541
 DB 168 LysLysArgLysLys-----ThrThrAsnValPheArgAla 179

RESULT 13
 ID 08BPD6 PRELIMINARY; PRT; 227 AA.
 AC 08BPD6;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Placenta specific homeobox 1.

GN PSX1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Placenta, and Extraembryonic tissue;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The PANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK076186; BAC36240.1; -.
 DR MGD; MGI:102888; Psx1.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; P:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 DR SEQUENCE 227 AA; 25544 MW; 0FEFF42ADB5F68F04 CRC64;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
8	53e-09	227	55
190	0.00	55	22
46	0.39	22	57
33	1.3	57	32
18	10	32	5

US-09-867-753-1 (1-577) x Q8BPD6 (1-227)

QY 104 GGGGAGCATCAAGCGCAGAGGCCATGTGGCCAGAGACTCCAGGCTCATGGTAA 163
 DB 54 GlyGluPheAspGlnGlyGluGluAlaGlnGlyGluValAlaGlyGlyGluGlnAlaGln 73
 QY 164 ATGAACCT-----GAGGCGGTGTAACCAAGAGCGGATG 202
 DB 74 GluGluProAlaProLeuSerProAlaGlnGluAlaThrGlyGlyGluGluGluGlu 93
 QY 203 AACCGGATGGGGGATG----- 220
 DB 94 AsnLysGlnGlyGluMetGlnGlyArgHisAlaGlyAspGlyAlaSerSerSerGluAsp 113
 QY 221 -----ATCCCGAGAGGCGCGGTGGAAC-----CAGAGCCTCGGCGAGCGCGAG 268
 DB 114 AspSerIleLeuGlnGluGlyGlyGluMetIleAspGlnGlnProProGlnGlnGluAla 133
 QY 269 CCCCCCGGAGAGCGCGCCAGCGCGCATGAGAGGCTCCGAGCCGAGAAC----- 322
 DB 134 AlaSerProAsp-----SerIleArgAsnProHisValLeuAsnArgLeu 148
 QY 323 ATGACGCGCAAGAACTCGGCGCAGAAAGTTGACGCTGTTCGAGTGAAGAGGAGTGA 382
 DB 149 AlaGlnLeuArgTyrArgArgThrArgPheThrHisPheGlnLeuHisAspLeuGluArg 168
 QY 383 GTTTCGACACATCAATACCTGATGTGCCCAAGAGAGGAATCTGCGAAAACTTA 442
 DB 169 LeuPheGlnGluThrArgTyrProSerLeuAlaGlyArgAspLeuAlaArgTyrPheMet 188
 QY 443 GGTGTGACTGAAGACAAAGTGGGTTGTTTAAAGATTAAGAGGCGCAGATGAGGCGA 502
 DB 189 GlyValAspGluCysAspValGlnAsnTyrPheArgMetArgArgAlaLeuPheGlnArg 208
 QY 503 CATCAGAGGAATTAATG 520
 DB 209 AsnArgArgValLeuMet 214

RESULT 14

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreanek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Gang N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jatali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA Mcintosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Plitman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter C.J., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RA Mista S., Crosby M.A., Mathews B.B., Bayraktaroglu I., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celisner S.E.,
 RA Clamp N., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seattle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RA FlyBase; (SEP-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS 1 PAIRED BOX DOMAIN.

DR EMBL: AY069177; AAL39322.1; -;
 DR EMBL: AB003541; AAF49919.2; -;
 DR FlyBase: FBgn0036285; toe.

DR GO: GO:0005634; C:nucleus; IEA.

DR GO: GO:0003700; P:transcription factor activity; IEA.

DR GO: GO:0007275; P:development; IEA.

DR GO: GO:0006355; P:regulation of transcription; IEA.

DR GO: GO:0006350; P:transcription; IEA.

DR InterPro: IPR001356; Homeobox.

DR InterPro: IPR001523; Paired_box.

DR Pfam: PF000046; homeobox; 1.

DR Pfam: PF00292; PAX; 1.

DR PRINTS: PR00027; PAIREDBOX.

DR ProDom: PD000010; Homeobox; 1.

DR SMART: SM00389; HOX; 1.

DR SMART: SM00351; PAX; 1.

DR PROSITE: PS00027; HOMEBOX_1; 1.

DR PROSITE: PS00027; HOMEBOX_2; 1.

DR DNA-binding; Developmental protein; Homeobox; Nuclear protein;

DR Paired box; Transcription regulation.

DR KW Paired box; Transcription; 306292EAF48B32B CRC64;

DR SO SEQUENCE 640 AA; 66525 MW; 306292EAF48B32B CRC64;

Alignment Scores:

Pred. No.:	9 34e-09	Length:	640
Score:	130.00	Matches:	55
Percent Similarity:	43.86%	Conservative:	20
Best Local Similarity:	32.16%	Mismatches:	60
Query Match:	18.10%	Indels:	36
DB:	5	Gaps:	7

US-09-867-753-1 (1-577) x Q8T0M4 (1-640)

QY 92 AACCTCAGCTGGGGGAGCATCAAGCCAGAGGCGCATGTGGCCAGAGCTCCAGGC 151
 |||||
 DB 279 ThrProProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 298
 |||||
 QY 152 CTCATG-----GTAATATG-----AACCT-----GAGGCGGCTGTG 184
 |||||
 DB 299 SerSerTyrglySerAspIysAspIysAsnMetSerSerAsnProAsnSerSerAsnSerAnthr 318
 |||||

QY 185 AACACAGAAACGGCATGACCGCGATGGCGCG----- 217
 |||||
 DB 319 ThrHisSerAsnGlyHisAsnThrAsnSerGlyCyseGlyAspSerSerIaGly 338
 |||||
 QY 218 -----ATGATCCCGAGGGCGGCGGTGGAAAAACAGAGACCT 253
 |||||
 DB 339 SerGlyArgLeuSerLeuProAlaLeuSerProAspSerGlySer----- 353
 |||||
 QY 254 CGGACAGACCGCCAGCGCCCGGAGAGACCGGCGCCAGCGCCATGGAGGGTCCGACG 313
 |||||
 DB 354 ArgAspSerArgSerProAspAlaAspAlaAsnArgMetIleAspIleGluGlyIuAsp 373
 |||||
 QY 314 CCCGAGAACATG-----CAGCCAGAACTCCG-----CGACGAAAGTTACGCTG 358
 |||||
 DB 374 SerGluSerGlnAspSerAspGlnProIysPheArgAlaGlnArgThrThrPheSerPro 393
 |||||
 QY 359 TTGACGTGAGAGAGCTGAAAGTGTTCGACACACTCAATACCTGATGTCGCCACA 418
 |||||
 DB 394 GluGlnLeuAspGluLeuGluIuysGluPheAspIysSerHisIetyrProCyseValAsnThr 413
 |||||
 QY 419 AGAAGGAACTTGGCCGAAACTTAGTGCTGACTGAAGCAAGTGGCGGTTGTTAAG 478
 |||||
 DB 414 ArgGluIuysLeuAlaIaArgThrAlaLeuSerGluAlaArgValGlnValItrPheSer 433
 |||||
 QY 479 AATAAAGGCGCCAGATGTAGCGCATCAGAGA 511
 |||||
 DB 434 AsnArgArgAlaIuysTrpArgArgHisGlnArg 444
 |||||

Search completed: June 16, 2004, 19:02:26
 Job time : 52 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 16, 2004, 18:56:45 ; Search time 20.5 Seconds
(without alignments) 5414.879 Million cell updates/sec

Title: US-09-867-753-1

Perfect score: 1050
Sequence: 1 tccacatcagcgcgtccag.....tctacatcgtcgtgacacag 577

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+np.model -DB=xlp
-Q=/cgn2.1/USPTO_epool_p/US09867753/runat_16062004_163636_27806/app_query.fasta_1.775
-DB=PIR_78 -QFMT=fastan -SUFFIX=pir -MIMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NCOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09867753_QCEN_1_1_38_0runat_16062004_163636_27806 -NCPU=6 -ICPU=3
-NO MMAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	196	18.7	384	2	A46403 transcription fact
2	189.5	18.0	399	2	A46403 paired-type homeod
3	189	18.0	252	1	A54677 homeotic protein g
4	188	17.9	245	1	A47539 homeotic protein g
5	187	17.8	256	1	A42768 homeotic protein g
6	184.5	17.6	217	2	I48902 homeobox protein g
7	184.5	17.6	245	2	S26076 homeobox protein k
8	181	17.2	419	1	S70617 homeotic protein k
9	180	17.1	333	1	JW0097 bicoid-related hom
10	179.5	17.1	326	2	A45452 transcription fact
11	179.5	17.1	362	2	T24046 hypothetical prote
12	178	17.0	185	2	A55882 homeobox protein H
13	176.5	16.8	290	2	I49265 paired box transcr
14	176	16.8	245	2	I51226 homeodomain protei

15	176	16.8	247	1	JC6540 placenta specific-C
16	176	16.8	280	2	I48713 Phox2 homeodomain
17	176	16.8	370	2	UC6130 paired box transcr
18	176	16.8	479	1	S15031 paired box transcr
19	176	16.8	612	2	A54282 reversed polarity
20	173	16.5	80	2	A26332 homeotic protein B
21	173	16.5	346	2	I48185 gene alix protein
22	173	16.5	427	1	A43698 paired box transcr
23	172.5	16.4	613	1	A26062 paired box segment
24	172.5	16.4	798	2	S20881 homeotic protein p
25	172	16.4	449	1	B43698 paired box transcr
26	172	16.4	520	2	S78502 paired box transcr
27	171.5	16.3	243	1	B47668 homeotic protein g
28	171.5	16.3	243	2	I51424 homeotic protein g
29	171.5	16.3	363	2	UC7750 CVC domain-contain
30	168.5	16.0	242	2	A43904 homeotic protein G
31	168.5	16.0	278	2	A56570 homeobox protein D
32	167.5	16.0	283	2	I50112 Dlx4 homeodomain p
33	167	15.9	288	2	J80659 homeotic protein H
34	167	15.9	319	2	T18786 hypothetical prote
35	167	15.9	326	2	A47523 cartilage homeopro
36	166.5	15.9	284	2	S60250 mab-18 protein (tr
37	166	15.8	296	2	S60251 paired type homeob
38	166	15.8	314	2	JC5273 homeodomain protei
39	166	15.8	318	2	S52424 homeodomain protei
40	166	15.8	319	2	T43635 homeodomain protei
41	166	15.8	373	2	A47234 homeobox protein H
42	165.5	15.8	288	2	S27842 homeotic protein s
43	165.5	15.8	671	2	A35912 homeotic protein o
44	164	15.6	491	1	S60252 paired box transcr
45	163.5	15.6	298	2	I51410 transcription fact

ALIGNMENTS

RESULT 1
A46403
transcription factor with prd-type homeo domain and Pro/Gln-rich domain al - fruit fly (I
C/species: Drosophila melanogaster
C/date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #ext_change 15-Oct-1999
C/accession: A46403
R/Schneitz, K.; Spielmann, P.; Noll, M.
Genes Dev. 7, 114-129, 1993
A>Title: Molecular genetics of arisaeless, a prd-type homeo box gene involved in the mori
A/Reference number: A46403; MUID:93138380; PMID:8093690
A/Accession: A46403
A/Status: preliminary
A/Molecule type: nucleic acid
A/Residues: 1-384 <SCH>
A/Note: sequence extracted from NCBI backbone (NCBIN:123442, NCBIPI:123444)
C/superfamily: unassigned homeobox proteins; homeobox homology
C/keywords: DNA binding; homeobox; nucleus; transcription regulation
F/86-142/Domain: homeobox homology <HOX>

Alignment Scores:

Pred. No.: 6.19e-09 Length: 384
Score: 196.00 Matches: 47
Percent Similarity: 55.17% Conservative: 17
Best Local Similarity: 40.52% Mismatches: 25
Query Match: 18.67% Indels: 27
DB: 2 Gaps: 3

US-09-867-753-1 (1-577) x A46403 (1-384)
QY 244 CCAGAGCCTCGGACGACGACGACCCCGCC-----GGA 279
DB 31 ProGlySerSerAlaAlaSerAlaGlyAlaAlaLeuThrValSerMetSerValSerGly 50
QY 280 GAGCGCGCCGACGCGCGCCATGAGGCTCCGA----- 312
DB 51 GlyAlaProSerGlyAlaSerGlyAlaSerGlyGlyThrAsnSerProValSerAspGly 70
QY 313 -----GCCGAGACATGACGACGACGACGACCTCGG-----CGCAC 345

```

Db      71 AenserAspCysegluAlaaspGluTyraIa-ProlysaRglYsGlnaRgArGtyrArgth 90
QY      346 GAAGTTCAGGCTTCCTGAGTGTGAGAGCTGGAAGTGTTCCTCCAGACATCAATACCC 405
Db      90 rThrPheThrSerPheGlnLeuGlnGlnLeuGlnGlnValPsaIaPheSerArgThrIstyr 110
QY      406 TGAATGCCACACAGAGGAACTTCCGAAAACCTTAGGTGTGACTGAACAGCAAGTGGC 465
Db      110 oAspValPheThrArgGlnGlnLeuAlaMetCysIleGlyLeuThrGlnAlaArgIleG 130
QY      466 GGTTCGTTAAGATPAAAGGCGCAGATGATGCGCAATCAGAGA 511
Db      130 nValTrpPheGlnAsnArgArGAlaIalysTrpArgIysGlnGlnIlys 145

```

RESULT 2

paired-type homeodomain protein, Alx-4 - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 31-Mar-2000
 C/Accession: J06522
 R:Ou, S.; Li, L.; Wisdom, R.
 Gene 203, 217-223, 1997
 A>Title: Alx-4: cDNA cloning and characterization of a novel paired-type homeodomain pro
 A:Reference number: J06522; M0ID:98086222; PMID:9426253
 A:Contents: Embryo
 A:Accession: J06522
 A:Molecule type: mRNA
 A:Residues: 1-339 <OUA>
 A:Cross-references: GB:AF001465; NID:g2352265; PIND:AAC39943.1; PID:g2352266
 C/Comment: This protein belongs to the family of paired-type homeodomain proteins, it p
 C/Genetics:
 A:Gene: Alx-4
 C:Superfamily: homeotic protein Hox B3; homeobox homology
 C:Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation
 F:203-259/Domain: homeobox homology <HOX>

Alignment Scores:

Pred. No.:	2,24e-08	Length:	399
Score:	189.50	Matches:	55
Percent Similarity:	42.08%	Conservative:	22
Best Local Similarity:	30.05%	Mismatches:	49
Query Match:	18.05%	Indels:	57
DB:	2	Gaps:	7

US-09-867-753-1 (1-577) x J06522 (1-399)

```

QY      89 CCCACACCTTAG-----CTGGGGGCGACATCAAGCCGAGAGCCATGTTGCCAA 139
Db      95 ProThrProGlnProProAlaProProAlaProProAlaHisLeuTyrlaGlnArg 114
QY      140 GGAGCTCCAGGCTCATGGTAATATGAACCTGAGGGCGGTGTGAACACGAGAACGGC 199
Db      115 GlyAla-----CysIsthrProProAspIysleuIys----- 126
QY      200 ATGAACCGGATGCGGATGATCCCGAGAGGGCGGTGAAACAGAGACCTCGCAG 259
Db      127 -----LeuGlnGlnGlySerGlyIylHisAlaIalalaLeuGln 139
QY      260 CAGCCG-----CAGCCCCCGCGGAGAG 293
Db      140 ValProCysTyraIalysGlnSerAsnLeuGlyGlnProGlnLeuProProAspSerGlu 159
QY      284 CCGGCG-----CAGGCGGCGATGAGGGGTCCGCGAG 313
Db      160 ProValGlyMetAspAsnSerTyrlaSerValIysGlnThrGlyAlaIalysGlnProGln 179
QY      314 -----CCCGAGACATGACAG----- 328
Db      180 AspArgAlaSerAlaGlnIleProSerProLeuGlnIlyThrAspSerGlnSerAsnIys 199
QY      329 -----CCACGAACCTGGCGCGACGAAGTTACCGCTGTTGAGAGGCTGGAAGT 382

```

```

Db      200 GlyIysIysArgArgAsnArgThrThrPheThrSerTyrlaGlnLeuGlnIleuGlnIlys 219
QY      383 GTTTCGACACACATCAATACCTGATGTGCCACACAGAGGAACTTCCGAAAACTTA 442
Db      220 ValPheGlnIlyThrHisIsthrProAspValTyraIalysGlnGlnLeuAlaMetArgThr 239
QY      443 GGTGTGACTGAAGACAGACATGCGGGTGTGTTAAGATPAAAGGCGCAGATGTAGCGCA 502
Db      240 AspLeuThrGlnAlaArgValGlnIalTrpPheGlnAsnArgArGAlaIalysTrpArgIys 259
QY      503 CATCAGAGA 511
Db      260 ArgGlnArg 262

```

RESULT 3

homeotic protein goosecoid (validated) - human
 A:54677
 C/Species: Homo sapiens (man)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000
 C/Accession: A54677
 R:Blum, M.; De Robertis, E.M.; Kojis, T.; Heinzmann, C.; Klisak, I.; Geiseler, D.; Spark
 Genomics 21, 388-393, 1994
 A>Title: Molecular cloning of the human homeobox gene goosecoid (GSC) and mapping of the
 A:Reference number: A54677; M0ID:94375063; PMID:7916327
 A:Accession: A54677
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-252 <BLU>
 C/Genetics:
 A:Gene: GDB:GSC
 A:Cross-references: GDB:251683; OMIM:138890
 A:Map position: 14q32.1-14q32.1
 C:Superfamily: homeotic protein goosecoid; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:156-212/Domain: homeobox homology <HOX>

Alignment Scores:

Pred. No.:	2,46e-08	Length:	252
Score:	189.00	Matches:	62
Percent Similarity:	40.87%	Conservative:	23
Best Local Similarity:	29.81%	Mismatches:	65
Query Match:	18.00%	Indels:	58
DB:	1	Gaps:	6

US-09-867-753-1 (1-577) x A54677 (1-252)

```

QY      5 ACATCGAGGCTCCGAGCCATGGGGGCTGCTGTCACAGACACCGTGTCTACTGCTG 64
Db      27 ThrAlaAlaIalProValIalPheProAlaLeuHisGlyAspSerLeuTy 43
QY      65 AGTGTATACAGGTAAAAATAAGCCCACTCACTGAGGGGCAAGCATCAAGCGCAGAA 124
Db      44 -----GlyGlyIalaserAspTy 50
QY      125 GGCATGTT-----GGCAAGAGCTCCAGGCTC----- 154
Db      51 GlyAlaPheTyrlaArgProValAlaProGlyAlaGlyLeuProAlaIalValThrGly 70
QY      155 -----ATGGGTAAATATGAACCTGAGGGCGGTGTGAACACAGAGAAC----- 196
Db      71 SerArgLeuGlyTyrlaAsnAsnTyrlaTyrlaGlnLeuHisValGlnAlaIalProVal 90
QY      197 GGCATGAACCGGATGCGGCGCATGATCCCGAGAGGGCGGCGGTGAAACAGAGAGCTCGG 256
Db      91 GlyProAlaCysCysGlyIalValaIalProProLeuGlnIal-----GlnIlyCysSer 107
QY      257 CAGACCGCGACGCCCCCGGAGAGCGCGCCAGCGGCGCATGAGAGGATCCGACGCC 316
Db      108 CysValProThrProProGlyTyrlaGlnGlyProGlySerValLeuValSerProValPro 127
QY      317 GAGAAATGACAGCA----- 331
Db      128 HisGlnMetLeuProTyrlaMetAsnValGlyThrLeuSerArgThrIleuGlnIleuLeu 147

```

QY 332 -----CGAATCTGGCGGCGAGGATGACGCTGTTGAGAGG 367
DB 148 AengInLeuH1sCyArGrArgLySarGArgh1sArGTh1lPheThrsPgluInLeu 167
QY 368 GAGGAGCTGGAAAGTGTTCGACACACTCAATACCTGATGTGGCCCAAGAAAGGAA 427
DB 168 GluAlaLeuGlnAlaLeuPheGlnGluThrLySyrProAspValGlyThrsArgInGlu 187
QY 428 CTTCGCCGAAACTTGAAGTGTGACTGAGACCAAGTCGGGTTTGGTTTAAGATAAAG 487
DB 188 LeuAlaArgLySValH1sLeuArgGlnGluLySValGluValTrpPheLySAsnArg 207
QY 488 GCCAGATGTAGCGACATCAGAGA 511
DB 208 AlaLySTrpArgGlnLySarG 215
RESULT 4
A47539
homeotic protein gooseoid [similarity] - chicken
C/Species: Gallus gallus (chicken)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000
C/Accession: A47539
R:Ipizua-Belmonte, J.C.; De Robertis, E.M.; Storey, K.G.; Stern, C.D.
Cell 74, 645-659, 1993
A/Title: The homeobox gene 'gooseoid' and the origin of organizer cells in the early chick
A/Reference number: A47539; MUID:93364981; PMID:7916659
A/Accession: A47539
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-245 <12P>
A/Cross-references: EMBL:X70471; NID:9402580; PIND:CAA49897.1; PID:9402581
C/Superfamily: homeotic protein gooseoid; homeobox homology
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:151-207/Domain: homeobox homology <HOX>
Alignment Scores:
Pred. No.: 36-08 Length: 245
Score: 188.00 Matches: 57
Percent Similarity: 40.728 Conservative: 22
Best Local Similarity: 29.388 Mismatches: 51
Query Match: 17.904 Indels: 64
DB: 1 Gaps: 6
US-09-867-753-1 (1-577) x A47539 (1-245)
QY 2 CCAACATCAGGCGCTCCAGCCATGCGCGCTG---CTGCTCCAGCAGACCGTGTTCAC 58
DB 57 ProGlySerAlaLeuProAlaValAlaGlySerArgLeuGlyTyrsAsnSerTyrrTyrr 76
QY 59 TCCCTGAGTGTATACCAAGTAAATAAGCCCACTCAAGTGGGGCAGCATCAAGC 118
DB 77 -----GlyInLeuH1sValAlaThrSer 84
QY 119 GCAGAAAGCCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTAAATGAACCTGAGGC 178
DB 85 ProValGly-----ProSerCysCysGlyAlaValAlaProProLeuGly 98
QY 179 GGTGTGAACCAAGAACCGCATGAACCGCATGTGGCGGATGATCCCGAGGGCGCGGT 238
DB 99 Ala-----GlnGlnCysSerCysValAlaProProAlaGly---- 109
QY 239 GGAACCAAGAGCCTTGGCAGCAGCGCCAGCCCGCGAGAGAGCCGCGCCAGGGCGC 298
DB 110 -----TyrluGlnValAlaGlySerVal 116
QY 299 ATGAGAGGTCCGAGCCGAGAAATGATGAGCCA----- 331
DB 117 LeuMetSerProValProH1sGlnMetLeuProTyMetAsnValGlyThrsLeuSerArg 136
QY 332 -----CGAATCTGGCGGCGAG 349
DB 137 ThrGluLeuGlnLeuLeuAengInLeuH1sCyArGrArgLySarGArgh1sArGThr1le 156

QY 350 TTCACTGTTGTCAGGTGAGAGAGCTGAAAGTGTTCGACACACTCAATACCTGAT 409
DB 157 PheThrArgGlnGlnLeuGlnAlaLeuGlnLeuH1sLeuPheGlnGlnThrLySyrTrpAsp 176
QY 410 GTGCCCAAGAGAGGAACTTGCCGAAACTTAGGTGTGACTGAAGACAAAGTGGCGGT 469
DB 177 ValGlyThrsArgGlnGlnLeuAlaArgLySValH1sLeuArgGlnGlnLySValGluVal 196
QY 470 TCGTTTAAGATAAAGCGCCGATGTAGCGACATCAGAGA 511
DB 197 TrpPheLySAsnArgAlaLySTrpArgGlnLySarG 210
RESULT 5
A42768
homeotic protein gooseoid [similarity] - mouse
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000
C/Accession: A42768
R:Blum, M.; Gaunt, S.J.; Cho, K.W.Y.; Steinbeisser, H.; Blumberg, B.; Bltner, D.; De Rot
Cell 69, 1097-1106, 1992
A/Title: Gastrulation in the mouse: the role of the homeobox gene gooseoid.
A/Reference number: A42768; MUID:92315328; PMID:1352187
A/Accession: A42768
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-256 <BLU>
A/Cross-references: GB:M85271; NID:9193895; PIND:AAA7826.1; PID:9193896
A/Note: sequence extracted from NCBI backbone (NCBI:P108110)
C/Superfamily: homeotic protein gooseoid; homeobox homology
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:161-217/Domain: homeobox homology <HOX>
Alignment Scores:
Pred. No.: 3,666-08 Length: 256
Score: 187.00 Matches: 59
Percent Similarity: 39.53% Conservative: 26
Best Local Similarity: 27.44% Mismatches: 64
Query Match: 17.81% Indels: 66
DB: 1 Gaps: 6
US-09-867-753-1 (1-577) x A42768 (1-256)
QY 2 CCAACATCAGGCGCTCCAGCCATGCGCGCTGCTCCAGCAGACCGTGTTCATCTG 61
DB 27 ProSerAlaAlaAlaProValAlaPheProAlaLeuH1sGlyAspSerLeuTyrr----- 44
QY 62 CTGAGTGTATACCAAGTAAATAAGCCCACTCAAGTGGGGCAGCATCAAGCGCA 121
DB 45 -----GlyAlaGlyGlyGlyThr 50
QY 122 GAAGGCCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTAAATGAACCTGAGGC--- 178
DB 51 SerSerAspTyrglyAlaPheTyrrProArgProValAla-----ProGlyGlyAla 67
QY 179 GGTGTGAACCAAGAGAGCGCATGAACCGCATGTC----- 214
DB 68 GlyLeuProAlaAlaValAlaGlySerSerArgLeuGlyTyrsAsnSerTyrrPheTyrglyGln 87
QY 215 -----GGCATGATCCCGAGGGCGC 235
DB 88 LeuH1sValGlnAlaAlaProValGlyProAlaCysCysGlyAlaValAlaProProLeuGly 107
QY 236 GGTGAAACCAAGAGCCTTGGCAGCAGCGCCGCGCCGCGAGAGAGCCGCGCCAGGG 295
DB 108 Ala-----GlnGlnCysSerCysValProThrProGlyTyrglyGlnProGlySer 125
QY 296 GCATGAGAGGTCCGAGCCGAGAAATGATGAGCCA----- 331
DB 126 ValLeuValSerProValProH1sGlnMetLeuProTyMetAsnValGlyThrsLeuSer 145
QY 332 -----CGAATCTGGCGGCGAG 346

DB 146 ArgThGluLeuGlnLeuLeuAenGlnLeuHisCysArgArgLysArgRghIsArgThr 165
 QY 347 AAGTTCACGCTGTTGGAGTGAGGAGCTGGAAGGTTTCCGACACACTCAATACCT 406
 DB 166 TlePheThraSpoluglnLeuGlnAlaLeuGlnAenPheGlnGlnThrlsTYrPro 185
 QY 407 GATGTCACCAAGAAAGGAACTTGCAGAAACTTAAGTGTGACTGAAGCAAAAGTCCG 466
 DB 186 AspValGlyThrArgGlnGlnLeuAlaArgLysValHisLeuArgGlnGlnLysValGlu 205
 QY 467 GTTGTGTTTAAGAAATAAAGGCGCAGATGATGAGCGACATCAGCA 511
 DB 206 ValTrpPheLysAsnArgAlaLysTrpArgArgGlnLysArg 220

RESULT 6

homeobox protein Pmx - mouse
 N/Alternate names: homeotic protein K-2b
 C/Species: Mus musculus (house mouse)
 C/Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Sep-1999
 C/Accession: I48902; I53118; S26077
 R/Kern, M.J.; Argao, E.A.; Birkenmeier, E.H.; Rowe, L.B.; Potter, S.S.
 Genomics 19, 334-340, 1994
 A/Title: Genomic organization and chromosome localization of the murine homeobox gene Pmx
 A/Reference number: I48902; PMID:94245205; PMID:7910581
 A/Accession: I48902
 A/Status: Preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-217 <RES>
 A/Cross-references: EMBL:U03873; NID:g460124; PIDD:AA52139.1; PID:g460125
 R/Kern, M.J.; Lilly, B.; Bryson, L.J.; Wang, Y.; Saesoon, D.A.; Olson, E.N.
 Development 115, 1087-1101, 1992
 A/Title: MHOX: a mesodermally restricted homeodomain protein that binds an essential site
 A/Reference number: I53118; PMID:93083424; PMID:1360403
 A/Accession: I53118
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-217 <RES>
 A/Cross-references: GB:U06902; NID:g199583; PIDD:AAA3672.1; PID:g199584
 R/Kern, M.J.; Witte, D.P.; Valerius, M.T.; Aronow, B.J.; Potter, S.S.
 Nucleic Acids Res. 20, 5189-5195, 1992
 A/Title: A novel murine homeobox gene isolated by a tissue specific PCR cloning strategy
 A/Reference number: S26076; PMID:93027261; PMID:1383943
 A/Accession: S26077
 A/Molecule type: mRNA
 A/Residues: 1-217 <XER>
 A/Cross-references: EMBL:X59726
 A/Experimental source: fetal heart
 C/Genetics:
 A/Gene: Pmx
 C/Superfamily: unassigned homeobox proteins; homeobox homology
 C/Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regulation
 F/95-151/Domain: homeobox homology <Hox>

Alignment Scores:

Pred. No.:	6e-08	Length:	217
Score:	184.50	Matches:	49
Percent Similarity:	44.97%	Conservative:	27
Best Local Similarity:	28.99%	Mismatches:	54
Query Match:	17.57%	Indels:	39
DB:	2	Gaps:	5

US-09-867-753-1 (1-577) x I48902 (1-217)

QY 113 TCAAGCGCAGAAAGGCAATGTTGGCAAGAGCTCCAGGCTCATGGGTATATGAACCT 172
 DB 2 ThnSerSerTyrlGlyHisValLeuGlnArgGlnProAlaLeuGlnGlnThrlsTYrPro 21
 QY 173 GAGGCGC-----GGTGAACCAAGCAAGGCGC 199
 DB 22 ProGlyAsnLeuAspThrLeuGlnAlaLysValAsnPheSerValSerHisLeuLeuAsp 41
 QY 200 ATGAACCGGATGCGCGGCGATGATCC-----GAGGCGCGC 235

DB 42 LeuGlnGlnAlaGlyAspMetValAlaAlaGlnAlaAspLysSerValGlyGlnAlaGly 61
 QY 236 -----GGTGAACCAAGGCGCTGCGGACGCGC 265
 DB 62 ArgSerLeuLeuGlnSerProGlyLeuThrSerGlySerAspThrProGlnGlnAsp--- 80
 QY 266 CAGCCCCCGCGAGAGAGCGCGCCAGCGCGCATAGAGAGGTCCGACGCCGAGAACATG 325
 DB 81 -----AsnAspGlnLeuAsnSerGlnGlnLysLysValArg 92
 QY 326 CAGCCACGAACTCGGCGCGCAAGATTACAGCTGTTCCAGCTGAGAGAGCTGGAAGTGT 385
 DB 93 LysGlnArgArgAsnArgThrThrPheAsnSerSerGlnLeuGlnAlaLeuGlnVal 112
 QY 386 TTCCGACACATCAATACCTGATGCGCCACAGAGGAACTTCCGAGAACTTAGGT 445
 DB 113 PheGlnArgThrHisTYrProAspAlaPheValArgGlnAspLeuAlaArgArgValAsn 132
 QY 446 GTGACTGAAGACAAAGTGGCGGCTTGTGTTTAAGAAATAAAGGCGCAGATGATGCGACAT 505
 DB 133 LeuThrGlnAlaArgValGlnValTrpPheGlnAsnArgArgAlaLysPheArgArgAsn 152
 QY 506 CAGAGAGAAATTAATGCTCCCAATCA 532
 DB 153 GluArg---AlaMetLeuAlaAsnLys 160

RESULT 7

homeotic protein K-2a - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 25-Feb-1994 #sequence_revision 01-Sep-1995 #text_change 24-Sep-1999
 C/Accession: S26076
 R/Kern, M.J.; Witte, D.P.; Valerius, M.T.; Aronow, B.J.; Potter, S.S.
 Nucleic Acids Res. 20, 5189-5195, 1992
 A/Title: A novel murine homeobox gene isolated by a tissue specific PCR cloning strategy.
 A/Reference number: S26076; PMID:93027261; PMID:1383943
 A/Accession: S26076
 A/Molecule type: mRNA
 A/Residues: 1-245 <XER>
 A/Cross-references: EMBL:X59725; NID:g51361; PIDD:CAA42410.1; PID:g51362
 A/Experimental source: fetal heart
 C/Superfamily: unassigned homeobox proteins; homeobox homology
 C/Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regulation
 F/95-151/Domain: homeobox homology <Hox>

Alignment Scores:

Pred. No.:	6e-08	Length:	245
Score:	184.50	Matches:	49
Percent Similarity:	44.97%	Conservative:	27
Best Local Similarity:	28.99%	Mismatches:	54
Query Match:	17.57%	Indels:	39
DB:	2	Gaps:	5

US-09-867-753-1 (1-577) x S26076 (1-245)

QY 113 TCAAGCGCAGAAAGGCAATGTTGGCAAGAGCTCCAGGCTCATGGGTATATGAACCT 172
 DB 2 ThnSerSerTyrlGlyHisValLeuGlnArgGlnProAlaLeuGlnGlnThrlsTYrPro 21
 QY 173 GAGGCGC-----GGTGAACCAAGCAAGGCGC 199
 DB 22 ProGlyAsnLeuAspThrLeuGlnAlaLysValAsnPheSerValSerHisLeuLeuAsp 41
 QY 200 ATGAACCGGATGCGCGGCGATGATCC-----GAGGCGCGC 235
 DB 42 LeuGlnGlnAlaGlyAspMetValAlaAlaGlnAlaAspLysSerValGlyGlnAlaGly 61
 QY 236 -----GGTGAACCAAGGCGCTGCGGACGCGC 265
 DB 62 ArgSerLeuLeuGlnSerProGlyLeuThrSerGlySerAspThrProGlnGlnAsp--- 80
 QY 266 CAGCCCCCGCGAGAGAGCGCGCCAGCGCGCATAGAGAGGTCCGACGCCGAGAACATG 325


```
DB -----AenAeRgLnLeuAenSerGluGluValuYbLybLybArg 92
QY 326 CAGCCACGAAGCTGGGGCAGCAAGTTTACAGCTTTGACAGGTGAGAGAGCTGGAAAGTGT 385
DB 93 LysGlnArgGlnArgGlnArgGlnArgGlnArgGlnArgGlnArgGlnArgGlnArgGlnArgGln 112
QY 386 TTCCGACACATCAATACCTGATGTGACCCACAAAGAGGAACTTGCCGAAACTTAGGT 445
DB 113 PheGluArgThrHisTyrProAlaPheValArgGluPheValAlaArgValaen 132
QY 446 GTGACTGAAGACAAAGTGGGGTTTGTTTAAAGATTAAGAGGACAGATGTAGCCACAT 505
DB 133 LeuThrGluAlaArgValGlnValTrrPheGlnAenArgGlnAlaYbPheArgAaen 152
QY 506 CAGAGAGAAATTAATGCTCCGCAATGAA 532
DB 153 GluArg---AlaMetLeuAlaAenLyb 160

RESULT 8
S70617
homoeotic protein gooseoid - fruit fly (Drosophila melanogaster)
N:Alternate names: homoeotic protein gooseoid
C:Species: Drosophila melanogaster
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 17-Nov-2000
C:Accession: S70617
R:Hahn, M.; Jaekle, H.
EMBO J. 15, 3077-3084, 1996
A:Title: Drosophila gooseoid participates in neural development but not in body axis fo
A:Reference number: S70617; PMID:96272167; PMID:8670808
A:Accession: S70617
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-419 <HAH>
A:Cross-references: EMBL:U52968; NID:g1399586; PIDN:AA017948.1; PID:g1399587
C:Genetics:
A:Gene: gsc
A:Cross-references: FlyBase:FBgn010323
A:Map position: 2
C:Function:
A:Description: plays a role in neurogenesis in post-gastrula Drosophila embryos
A>Note: not required for gastrulation like Xenopus gooseoid; expressed most strongly in
C:Superfamily: fruit fly homoeotic protein gooseoid; homoeobox homology
C:Keywords: DNA binding; homoeotic; nucleus; transcription regulation
F:287-343/Domain: homoeobox homology <HOK>

Alignment Scores:
Pred. No.: 1,21e-07 Length: 419
Score: 181.00 Matches: 48
Percent Similarity: 49.30% Conservative: 22
Best Local Similarity: 33.80% Mismatches: 42
Query Match: 17,24% Indels: 30
DB: 1 Gaps: 5

US-09-867-753-1 (1-577) x S70617 (1-419)
QY 107 GGAGATCAAGCCGAGAGCCATGTTGGCCAAAGAGCTCCAGGCTCATGGTAT--- 163
DB 230 AAlaAlaAlaGlnMetGlnAlaHisValSerGluAlaAlaAlaGluLeuSerGluHisGly 249
QY 164 -----ATGAAACCTGAGGCGGTGTGAACCCAGAGAAC---GGCATGAACCCCGAT 211
DB 250 HisLeuProHisLeuSerHisLeuGluHisLeuProHisLeuGluAlaHisHisHis 269
QY 212 GGC---GGCATGATCCCGAGGCGCGGTGAAACCCAGAGACCTTGCGAGAGCCGCGAG 268
DB 270 GluGlnHisHisLeuSerHisLeuGluHisGly----- 280
QY 269 CCCCCGCGAGGAGCGGCGCCAGGCGGCATGAGAGGTCGCGAGCCGAGAAATGATGAG 328
DB 281 ProProGluLyb-----Arg 285
QY 329 CCAAGAACTGGCGGAGAGAAATGATCAAGCTGTTGCAAGTGAGAGAGCTGGAAGTGTTC 388
```

```
DB 286 LysArgArgHisArgThrIlePheThrGluGlnLeuGluGlnLeuGluAlaThrPhe 305
QY 389 CAGACACATCAATACCTGATGTGACCCACAAAGAGGAACTTGCCGAAACTTAGGTG 448
DB 306 AspLysThrHisTyrProAspValValLeuArgGluGlnLeuAlaLeuLybAlaPhe 325
QY 449 ACTGAAGACAAAGTGGGGTTTGTTTAAAGATTAAGAGGACAGATGTAGCCACATGAG 508
DB 326 LysGluGluArgValGluValTrrPheLybAenArgValAlaYbTrrPheLybGlnLyb 345
QY 509 AGAGAA 514
DB 346 ArgGlu 347

RESULT 9
JW0097
bicoid-related homoeobox protein Ptx2 - chicken
C:Species: Gallus gallus (chicken)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Oct-1999
C:Accession: JW0097
R:St. Amand, T.R.; Ka, J.; Zhang, Y.; Hu, Y.; Baber, S.I.; Qiu, M.S.; Chen, Y.P.
Biochem. Biophys. Res. Commun. 247, 100-105, 1998
A:Title: Cloning and expression pattern of chicken Ptx2: A new component in the SHN sigr
A:Reference number: JW0097; PMID:98300291; PMID:9636662
A:Accession: JW0097
A:Molecule type: mRNA
A:Residues: 1-333 <STA>
A:Cross-references: GB:AF076640; NID:g3335642; PIDN:AAC27322.1; PID:g3335643
C:Comment: This protein is a new component in the Sonic hedgehog signaling pathway and p)
C:Genetics:
A:Gene: cPtx2
C:Superfamily: chicken bicoid-related homoeobox protein Ptx2; homoeobox homology
C:Keywords: DNA binding; homoeotic; nucleus; transcription regulation
F:102-158/Domain: homoeobox homology <HOK>

Alignment Scores:
Pred. No.: 1,47e-07 Length: 333
Score: 180.00 Matches: 48
Percent Similarity: 45.56% Conservative: 29
Best Local Similarity: 28.40% Mismatches: 40
Query Match: 17,14% Indels: 52
DB: 1 Gaps: 5

US-09-867-753-1 (1-577) x JW0097 (1-333)
QY 5 ACATCAGGCGCTCCAGCCATGCGCGCTTGCTGCCAGAGACCGTGTCTATGCTGTG 64
DB 45 ThrAlaLeuAlaProGluGlnAlaArgSerSerLeu---GluAlaAlaLybHisArgLeu 63
QY 65 AGTGTATACCAAGTAAATAAGCCCCACACTCAGCTGGGGGACATCAAGCGAGAA 124
DB 64 GluValHisThrIleSerAspThrSerSerProGlu---AlaAlaGlu 78
QY 125 GGCAATGTTGGCCAAAGAGCTCCAGGCTCATGGGTATATGAACCTTGAGGCGGTGTG 184
DB 79 LysGluLybSerGln----- 83
QY 185 AACCAAGAGAGCGCATGAAACCGAGTGGCGCATGATCCCGAGGCGCGCGGTGAAAC 244
DB 84 -----GlnGlyLysSerGluPheArgGly----- 91
QY 245 CAGAGACCTTGCGAGAGCGGAGAGCCCGCGCGAGAGACCGGCGGCGCATGAG 304
DB 92 -----ProGluAspProSer----- 96
QY 305 GGTCCGAGCCGAGAAATGCAATGCAAGCAAGTCCGCGGAGAAATGCAAGTTCAGG 364
DB 97 -----LysLybLysArgGlnArgGlnArgGlnArgGlnHisPheThrSerGlnGln 112
QY 365 GTGAGAGAGCTGGAAGTGTTCGAGACACTCAATACCTGATGTGTGCGCAAGAGAGG 424
DB 113 LeuGlnGluLeuGlnAlaThrPheGlnArgAenArgGlyProAspMetSerThrArgGlu 132
```

[illegible]

```

Oy      442 AGGTGTGACCTGGAAGCAAAATGCGGGTTTGGTTTAAGATTAAGAGGGCGCATGTAGCGC 501
Db      103 alyb/leu1nrg1a1aargValGlnValThrPheSerAsnArgAlaArgTTPargLy 123
Oy      502 ACAT-----CAGAGGAATTAATGCTCGCCCAATGACTA 535
Db      123 sgl1a1aGlyAlaAsnGlnLeuMetAlaPheAsnHisLeu 136

RESULT 11
T24046
hypothetical protein R08B4.2 - Caenorhabditis elegans
C|Species: Caenorhabditis elegans
C|Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Mar-2000
C|Accession: T24046
R|White, S.
submitted to the EMBL Data Library, November 1995
A|Reference number: Z19834
A|Accession: T24046
A|Status: preliminary; translated from GB/EMBL/DDBJ
A|Molecule type: DNA
A|Residues: 1-362 <MIL>
A|Cross-references: EMBL:Z68008; PIDN:CAA92001.1; GSPDB:GNO0028; CESP:R08B4.2
A|Experimental source: clone R08B4
C|Genetics:
A|Gene: CESP:R08B4.2
A|Map position: X
A|Intons: 43/3; 85/1; 147/2; 162/3; 222/1
C|Superfamily: unassigned homeobox proteins; homeobox homology
C|Keywords: DNA binding; homeobox; nucleu; transcription regulation

Alignment Scores:
Pred. No.:      1.62e-07      Length:      362
Score:          179.50      Matches:      39
Percent Similarity: 49.548      Conservative: 15
Best Local Similarity: 35.784      Mismatches:  28
Query Match:      17.104      Indels:      27
DB:              2          Gaps:          1

US-09-867-753-1 (1-577) x T24046 (1-362)

Oy      185 AACCCAGAGAACCGCATGAACCGCATGCGCGCATGATCCCGAGGGCGCGTGGAAAC 244
Db      96  AsnArgGluAsnIlySerProSerAspGlyThrAsnSerProAspAsnGlyLyArg 115
Oy      245 CAGAGAGCTCGGAGAGAGCGCGAGCCGCCCGCGAGAGCGCGGCCAGCGGCATGGAG 304
Db      116  LysGln-----117
Oy      305 GGTGCCGAGCCGAGAACATGATCGAGCCAGAACTCGCGCGCAGAGATTACAGCTTTGCAG 364
Db      118 -----ArgargGlyrArgThrThrPheSerAlaPheGln 128
Oy      365 GTGAGAGAGCTGGAAGATGTTTCCGACACACTCAATACCTGATGTGCCCCCAAGAGG 424
Db      129  LeuAspGlnLeuGlnIlyValPheAlaArgThrHisTyProAspValPheThrArgGlu 148
Oy      425 GAACCTGCGCAAAACTTAGTGTGACTGAACGCAAAAGTCCGGTTTGTTTAAGAAATAA 484
Db      149  GlnLeuAlaThrArgValGlnLeuThrGlnAlaArgValGlnValTyPheGlnAsnArg 168
Oy      485 AGGCGCAGATGTAGCGGACATCAGAGA 511
Db      169  ArgAlaIyeryrArgIlyGlnIlyArg 177

RESULT 12
A55882
homeobox protein Hsxl - mouse
N|Alternate names: homeotic protein HES-1
C|Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 24-Sep-1999
C|Accession: A55882; S35540; S49199
C|Thomas, P.O.; Johnson, B.V.; Rathjen, J.; Rathjen, P.D.

```

F:1-125/Domain:paired box homology (fragment) <PBH>
F:182-238/Domain: homeobox homology <HOX>

Alignment Scores:

Pred. No.:	2,936-07	Length:	290
Score:	176.50	Matches:	45
Percent Similarity:	43.28%	Conservative:	13
Best Local Similarity:	33.58%	Mismatches:	43
Query Match:	16.81%	Indels:	33
DB:	2	Gaps:	3

US-09-867-753-1 (1-577) x 149265 (1-290)

```

QY      173 GAGGCGGCTGTGACACGAGAACGGC----- 199
           |||||
Db      136 GUGUGAaplylYsGLUGlunapGLyGluYsLYsAlaLYshISerlleapGLyIle 155
           |||||

QY      200 ATGAACCCGATGCGGCGATGATCCCGAGGGCGGCGTGGAACACGAGCCTCGCAG 259
           |||||
Db      156 LeuGLyaplylYsGLYAsnArGLeuaapGLUGlySerdlYalGLuSerGLuPro----- 173

QY      260 CAGCCGACGCCCGCCGCGAGAGAGCGGCCGACGATGAGGTCGCGACGCCGAG 319
           |||||
Db      174 -----AspleuProleu 177

QY      320 AACATGACGCCACCAACTCGGCGACGAAAGTTACGCTGTTCAGGTGAGAGACTGGAA 379
           |||||
Db      178 LYsArGLyGLInArGLysSerArGLYrThrPherThAlaGLUGlunGLUGlunLeuGLu 197
           |||||

QY      380 AGTGTTCGCGACACATACCTGATGTCGCCACAGAAAGGAGAACTTGCCGAAC 439
           |||||
Db      198 LYsAlaapheGLuarThrHisTyrProapleIeTyrThrGLUGlunLeuAlaGLInArG 217
           |||||

QY      440 TTAGGTGTGACTGAAGACAAAGTCGGGTTGGTTTAAAGATTAAGAGGCCAGATGAG 499
           |||||
Db      218 ThrLYsleuThrGLuAlaArpheGLInValTrrPheSerasnArGLaArGLyTrrpArg 237
           |||||

QY      500 CGACATCAGAGAGAAATTAACTGCCCAATGAACCTACGTCT 541
           |||||
Db      238 LYsGLInaGLy-----AlaenGLInLeuAlaLa 247
           |||||

RESULT 14
151226
homeodomain protein - chicken
N:Alternate names: homeotic gene paired-related protein
C:Species: Gallus gallus (chicken)
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 17-Nov-2000
C:Accession: I51226; I50413
R:Kuratani, S.; Martin, J.F.; Wawerik, S.; Lilly, B.; Etchele, G.; Olson, E.N.
Dev. Biol. 161, 357-369, 1994
A:Title: The expression pattern of the chick homeobox gene gMHOX suggests a role in patte
A:Reference number: I51226; MUID:94148118; PMID:7906232
A:Accession: I51226
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-245 <KDX>
A:Cross-references: GB:S69088; NID:G545359; PIDN:AA29880.1; PID:G545360
R:Nohno, T.; Koyama, B.; Miyokai, F.; Taniguchi, S.; Ohuchi, H.; Saito, T.; Noji, S.
Dev. Biol. 158, 254-264, 1993
A:Title: The chicken homeobox gene related to Drosophila paired is predominantly expressed
A:Reference number: I50413; MUID:93321789; PMID:8101172
A:Accession: I50413
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-154,156-245 <NOH>
A:Cross-references: GB:D13433; NID:G222850; PIDN:BA02695.1; PID:G222851
C:Genetics:
A:Gene: gMHOX; Ptx-1
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:95-151/Domain: homeobox homology <HOX>
Alignment Scores:

```

